Training of the Beta wavelet networks by the frames theory: Application to face recognition

Mourad Zaied, Olfa Jemai and Chokri Ben Amar
(REsearch Group on Intelligent Machines (REGIM), University of Sfax, National Engineering School of Sfax, B.P. W, 3038, Sfax, Tunisia)

Abstract

A Wavelets Neural network is a hybrid classifier composed of a neuronal contraption and Wavelets as functions of activation. Our approach of face recognition is divided in two parts: the training phase and the recognition phase. The first consists in optimizing a Wavelets Neural network for every training picture face. A new technique of training of these wavelets networks which based on the frames theory is proposed as a remedy to the inconveniences of the classical training algorithms. The specificity of a BWNN to a face and the notion of SuperWavelet have been exploited to propose an approach of face recognition.

Finally, we have compared our method of recognition to other ones which are used for face recognition that are applied on the AT&T (ORL) and FERET faces basis. We reached a face recognition rate that exceeds 90% for two images per person in the training step.

Key Words- Orthogonal and bi-orthogonal wavelets, Wavelet Networks, frames, training, face recognition

1. Introduction

The recognition and the identification of faces play a fundamental role in our social interactions. It is based on our capacity to recognize people, it doesn't present any enormous difficulties for a human being but it constitutes for all computing system an extremely delicate situation. To invest in the domain of the face recognition is probably motivated by the multiplicity and the variety of applications fields (high security, telemonitoring and access control …). The works carrying out in this domain, under different conditions of lighting, facial expression and orientations can be classed in two different categories: The local approaches and the global approaches.

Among those approaches, the neural networks are presented on one hand by their capacity of approximation, which facilitates the learning of the faces to be recognized, and on the other hand by their property of classification shown in several domains of applications.

In spite of its advantage, a face recognition technique based on the neuronal networks present some problems notably in the learning algorithms used and during adding new persons to the recognition basis after training stage.

To remedy these problems, a solution based on the Wavelets networks (neural networks with wavelets as transfer functions) is proposed. The wavelets are also excellent approximators and signal analyzers. Their time-frequency analysis makes them an effective and innovative tool. Besides, their remarkable results in the domain of faces recognition (example the Gabor wavelet used with the method of EBGM) encourage their integration in such hybrid system.

In this paper, we introduce a new method to learning the WNN, which permit the direct calculation of the network connections weights by exploiting the frames theory. An approach of recognition based on the notion of SuperWavelet will be presented and detailed.

2. From the continuous wavelet transform (CWT) to frames

Daubechies, in [1] has mentioned that a function which satisfies the condition:

$$0 < C_{\psi} = 2\pi \int \frac{\left\|\hat{\psi}(w)\right\|^2}{\left\|w\right\|} dw < \infty \qquad (1)$$

is said admissible wavelet, $\hat{\psi}$ is the Fourier transform of the function ψ.

The equation (1) is often known as the admissibility condition.

For a function $f \in L^2(R)$ its continuous wavelet transform is given by:

$$w(a,b) = \frac{1}{\sqrt{|a|}} \int f(t)\psi(\frac{t-b}{a}) \text{ With } a \in R^*, b \in R \qquad (2)$$

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

$$f(x) = \frac{1}{C_{\psi}} \iint w(a,b) \frac{1}{\sqrt{|a|}} \psi(\frac{t-b}{a}) \frac{dadb}{a^2} \qquad (3)$$

2.1. The discrete wavelet transform (DWT)

It is known that the representation of the equation (2) is very redundant, so to analyze numeric signals, a sampling on a time-frequency grid, is necessary: The coefficients a and b will be sampled with the following

978-1-4244-3322-3/08/$25.00 ©2008 IEEE

manner: $a = a_0^m, b = nb_0 a_0^m$ with $a_0 > 1$ and $b_0 > 0$. So for a signal including a_0^j points we calculate only the coefficients:

$$w_{m,n}(f) = a_0^{-m/2} \int \psi(a_0^{-m} t - nb_0) f(t) dt \quad (4)$$
$$with\ m = 1,..., j\ \ n = 1,..., a_0^{j-m}$$

For $a_0 = 2, b_0 = 1$ the sampling is said dyadic. The gotten set of wavelets can constitute an orthogonal basis, in this case all function f can be written in a unique manner:

$$f(t) = \sum_{(m,n)\in S} w(m,n) \psi_{m,n}(t) = \sum_{(m,n)\in S} < \psi_{m,n}, f > \psi_{m,n}(t) \quad (5)$$

In a general case, we get wavelets that form frame, so we are led to write f according to the dual frame.

$$f(t) = \sum_{(m,n)\in S} < \tilde{\psi}_{m,n}, f > \psi_{m,n}(t) = \sum_{(m,n)\in S} < \psi_{m,n}, f > \tilde{\psi}_{m,n}(t) \quad (6)$$

3. The wavelet networks

The wavelets networks have been introduced by Zhang and Benveniste as a combination of the RBF neural network and the decomposition in wavelets.

3.1. Generalities

The Multilayered networks permit the representation of a non linear function by training while comparing their inputs and their outputs. This training is made while representing a non linear function by a combination of activation functions. The sigmoid function is often used as activation function.

Zhang and Benveniste [2] replaced the sigmoid function by an admissible wavelet and they demonstrated that a wavelet networks preserve the property of universal approximation of the RBF networks, a direct link exists between the weights of the network w_i and the wavelets coefficients and that a good approximation can be reached with a wavelet network of small size. The equation (3) gives the expression of a function f with two sums of all possible dilations and translations of the mother wavelet. Admitting now that we have only a finished number of n wavelets, we can consider the relation (7):

$$f(x) \approx \sum_{i=1}^{n} w_i \psi_i \quad (7)$$

As an approximation of the relation (5), the finished sum of the relation (7) is an approximation of the inverse transform. We can also consider it as a decomposition of the function f to a sum of w_i and wavelets.

To define a wavelet network, we start with taking a family of n wavelets $\Psi = \{\psi_1, ..., \psi_n\}$ with different parameters of scaling and translation that can be chosen arbitrarily at this point.

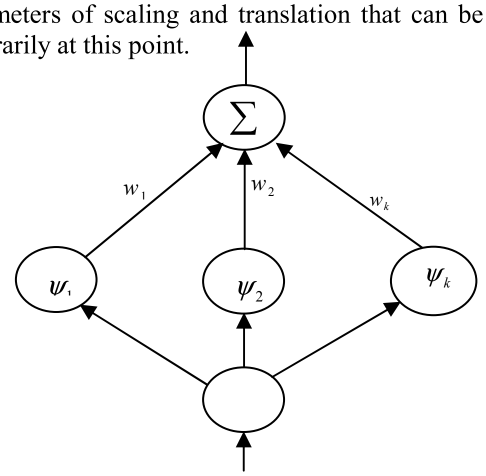

Figure 1. Wavelet network

Let's suppose that we have three wavelets ψ_1, ψ_2 and ψ_3 dilated and translated of only one mother wavelet:

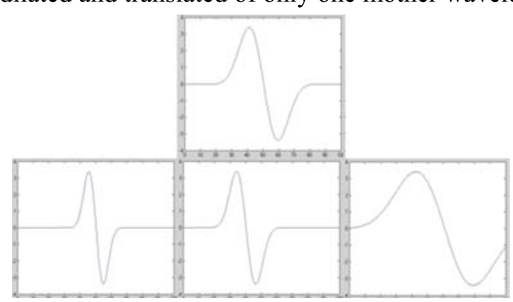

Figure 2. The mother wavelet on the top and its Three dilated and translated wavelets

A wavelet network constituted with these three wavelets can approximate a given signal f

3.2. Optimization of the wavelet networks by the frames theory

To find the optimal wavelet network of a given signal ***f***, the algorithm of the back propagation is generally used to minimize the energy function:

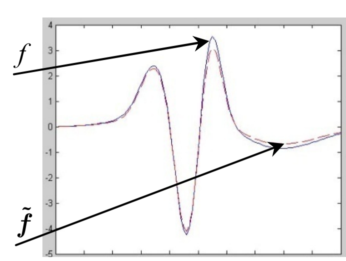

Figure 3. Approximation of a function f by a network of 3 wavelets

$$\tilde{f} = -\psi_{0,0.5} + 0.5\psi_{-10,0.75} + 0.25\psi_{10,2}$$

$$E = \min_{w_i \psi_i} \frac{1}{2}\left(f - \sum_{i=1}^{n} w_i \psi_i\right)^2 \quad (8)$$

Seen the inconveniences of this algorithm like the slowness or the convergence to local minima, we propose in this work a new method of training of the wavelet networks based on the frame theory.

3.2.1. Proposed learning algorithm

Step 1: Before start the learning, we prepare a library of candidate wavelets to be selected as activation functions of the wavelet network. This step is composed of the following:

1. Choosing the mother wavelet covering all the support of the signal to analyze
2. Building a library formed by the wavelets of the discrete wavelet transform using a dyadic sampling of the continues wavelet transform. These wavelets form a wavelet frame.
3. Select the lowest frequency wavelet of the library (the mother wavelet) as a first activation function (in this stage only one neuron is in the hidden layer).
4. Set as a stop learning condition an error E_{min} between the input and the output network or a number i of wavelet used for the learning or a number j of neurons in the hidden layer of the network.
5. The Learning is an incremental one. Each time we select the next wavelet of the library (the selection is sequential) and iterate the following steps:

Step 2: Calculate the dual basis formed by the activation wavelets of the hidden layer of the network and the new selected wavelet.

Step 3: If the selected wavelet ψ_n creates a basis (orthogonal or bi-orthogonal) with the (n-1) activation wavelets of the network, it's used as an activation function of a new neuron in the hidden layer; else it will update the (n-1) old weights of the network.

Step 4: Knowing the hidden layer wavelets and the connection weights, we can calculate the output of the network.

Step 5: If the error E_{min} or the number of wavelets used i or the number of neurons j are reached then it's the end of learning, else another wavelet of the library is selected and we return to Step 2

3.2.1.1. Creation of the library wavelet. To optimize a wavelet network, a sampling on a dyadic grid of continuous wavelet transform is used. We obtain a library wavelet candidate to join our wavelet network. In 1D wavelet dimension, this sampling gives in the first scale one wavelet (the mother wavelet). Every time that we climb a scale, the number of wavelet in this scale is multiplied by two.

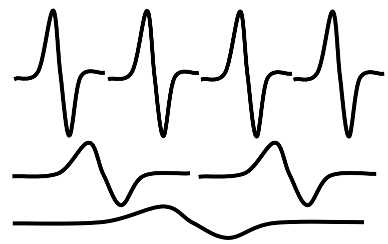

Figure 4. The wavelets of the three first scales

In 2D case the number of wavelet is multiplied by four between a scale and its successor.

3.2.1.2. How to verify if a chosen wavelet make basis with the others of the network? In every optimization stage, a wavelet from the library is introduced in the hidden layer of the network or used to up to date the weights. The wavelets of low frequency that allow a coarse approximation of the signal to be analyzed is introduced the first; those of the high frequency come to refine the approximated signal.

To know if a wavelet (n) will be an activation function of a new neuron, we must verify if it makes a basis with the (n-1) wavelets of the network. In other world it must be linearly independent with the mentioned (n-1) wavelets. For this, we define the Error function $E = f - \hat{f}$ with f the function to approximate and $\hat{f}$ the output of the network. At the beginning of the process of optimization $\hat{f} = 0$ and $E = f$. On the dyadic scale we take the first wavelet (the one of the lowest frequency). The output of the network is $\hat{f} = w_1\psi_1$ and $E = f - w_1\psi_1$. In a given stage $\hat{f} = \sum_i^{n-1} w_i\psi_i$ and $E = f - \sum_i^{n-1} w_i\psi_i$, the following wavelet on the sampling, that is make a basis with the (n-1) wavelets of the networks, going to join the network must verify the condition:

$$< D, \psi_n > = < f - \sum_i^{n-1} w_i \psi_i, \psi_n > \neq 0 \quad (9)$$

Demonstration: The wavelets network must form a basis, which means that these wavelets are linearly independent. The new wavelet must verify this condition and by consequence it must not belong to the space generated by the old wavelets: $\psi_n \notin < \psi_1, ..., \psi_{n-1} >$.

Admitting that $\psi_n \in < \psi_1, ..., \psi_{n-1} >$

We have therefore $< \psi_1, ..., \psi_{n-1} > = < \psi_1, ..., \psi_n >$

And in particular $(< \psi_1, ..., \psi_{n-1} >)^\perp = (< \psi_1, ..., \psi_n >)^\perp$

That means $f - \sum_{i}^{n-1} w_i \psi_i \in (<\psi_1,...,\psi_n>)^{\perp}$

What implies $<f - \sum_{i}^{n-1} w_i\psi_i, \psi_n> = 0$

That contradicts with the choice of ψ_n in the stage of optimization that must verify $<f - \sum_{i}^{n-1} w_i\psi_i, \psi_n> \neq 0$

So all the ψ_i are linearly independent and form a basis**.**

3.2.1.3. Calculation of weights. For orthogonal wavelets, the calculation of the weights connection to every stage is possible by projecting the signal to be analyzed on a wavelets family forming an orthogonal basis: $w_i = <f, \psi_i>$. The Beta wavelet is not necessarily orthogonal and can be also formed by wavelets that are not independent (frame). For a given family of wavelets it is not possible to calculate the weights by direct projection of the f function. We explain in this section that a simple calculation of the weights always remained possible even with non orthogonal wavelets.

Definition. Two families of scaling ψ_i and $\tilde{\psi}_i$ are said bi-orthogonal if for all i and j we have :

$$<\psi_i, \tilde{\psi}_j> = \delta_{i,j} \quad (10)$$

The wavelet ψ is said primal whereas the wavelets $\tilde{\psi}$ is said dual. If $\psi_i = \tilde{\psi}_i$, the family ψ_i constitutes an orthogonal basis.

if f a signal, ψ_i a family of wavelets who forms a bi-orthogonal or frame basis and $\tilde{\psi}_i$ the family of dual wavelets then some weights W_i exist such as : $f = \sum_i w_i \psi_i$. A weight can be calculated while exploiting the dual wavelet: $w_k = <f, \tilde{\psi}_k>$ (11)

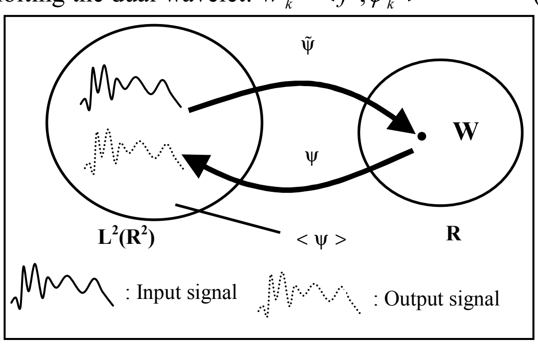

Figure 5. Computing weights and output Network process

3.2.1.4. How to calculate the dual wavelets family ?
At every stage of optimization process we are led to know the dual family of the wavelets forming our wavelet network.
The dual wavelets family is calculated by the formula [6]:

$$\tilde{\psi}_i = \sum_{j=1}^{N} (\Psi_{i,j})^{-1} \psi_j \quad with \quad \Psi_{i,j} = <\psi_i, \psi_j> \quad (12)$$

3.2.1.5. Optimization of the weights in the case of the frames. In the case of any frame (that doesn't form a basis), the values of the weights is not optimal since during the stage of optimization some wavelets have been separated. Let's suppose that we are at the second stage of optimization and that the first two wavelets are linearly independent, the output of the network is then:

$$\hat{f}_2 = w_1\psi_1 + w_2\psi_2 \quad (13)$$

Now suppose that at the third stage the wavelets are not linearly independent with the two firsts, the input signal projected on the dual family of the three wavelets leads to the approximation:

$$\hat{f}_3 = w_1\psi_1 + w_2\psi_2 + w_3\psi_3 \quad (14)$$

Since ψ_3 is depended of the two other wavelets, we can write:

$$\begin{aligned}\hat{f}_3 &= w_1\psi_1 + w_2\psi_2 + w_3(v_{3,1}\psi_1 + v_{3,2}\psi_2) \\ &= (w_1 + w_3 v_{3,1})\psi_1 + (w_2 + w_3 v_{3,2})\psi_2\end{aligned} \quad (15)$$

This lead to refine the coefficients of the second stage :

$$\hat{f}_2 = w'_1\psi_1 + w'_2\psi_2 \quad (16)$$

The $v_{i,j}$ are calculated by a projection on the dual basis of the family of the first two wavelets.
In general, at a stage n the weights of connections are updated by the formula:

$$\hat{f}_n = \sum_{i=1}^{m} (w_i + \sum_{j=m+1}^{n} w_j v_{j,i})\psi_i \quad (17)$$

With m the number of the wavelet of the network that are well linearly independent.

4. Faces recognition by Beta WN

4.1. The Beta wavelet

In our works, we have used the Beta wavelets as activation function. The Beta function is defined by [3]:

$$\beta(x; p, q, x_0, x_1) = \begin{cases} \left(\frac{x - x_0}{x_c - x_0}\right)^p \left(\frac{x_1 - x}{x_1 - x_c}\right)^q & if \ x \in]x_0, x_1[\\ 0 & otherwise \end{cases} \quad (18)$$

$$with \quad p, q, x_0 < x_1 \in \Re \quad and \quad x_c = \frac{px_1 + qx_0}{p + q}$$

We demonstrate in [3] [4] that all the derivatives of Beta function are admissible wavelets. We can obtain different wavelets when we modify the values of x0, x1, q and p parameters.

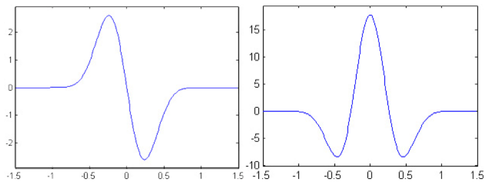

Figure 6. Beta 1 and Beta 2 1D Wavelets (First derivative and Second derivative of the Beta Function)

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).

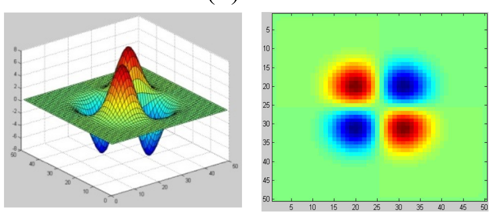

Figure 7. The 2D Beta1 wavelet and the associated filter

4.2. Training

Our goal is to recognize a face independently of any expression or any position. We proceed as follow: first a 2D wavelet network is optimized for every face, the functional parameters of every network (wavelet parameters and weight connections) are saved. We approximate only the faces area in the images. The next figure demonstrates the evolution of the output of a Beta wavelet network with the number of wavelets.

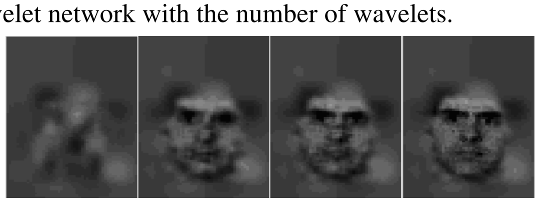

Figure 8. Variation of the approximation of an image of a face by a BWN2D using 10, 32, 64 and 175 wavelets

4.3. Recognition

4.3.1. Definition. Suppose (Ψ, W) is a Bêta Wavelet Network with $\Psi = (\psi_1 ... \psi_n)$, $W = (w_1 ... w_n)$. A Beta SuperWavelet Ψ is defined to be a linear combination of the Bêta wavelets ψ_i. The term SuperWavelet is defined for the first time by Szu [5]: $\Psi = \sum_{i=1}^{n} w_i \psi_i$ (19)

It is clear that the output of a wavelet network is a SuperWavelet. A SuperWavelet is a function that can be treated like a normal wavelet since it is the weighted sum of several wavelets. Szu associate the vector of the following parameters to this wavelet:

$$n = (c_x, c_y, \theta, s_x, s_y) \qquad (20)$$

With c_x and c_y the two parameters of translation, S_x and S_y the scaling parameters and θ the one of the rotation. V. Kruger [6] added a sixth parameter $S_{x,y}$ to make a dilatation in the two senses (x,y) at a time to assure all linear modifications of the SuperWavelet.

4.3.2. Recognition steps. Supposing that a person's g image is presented to the system to be recognized, the stages of recognition are the following:

Step 1: First the Beta SuperWavelet of every wavelet networks (Ψ_i, W_i) of the training basis is modified linearly so that the network approximates the image **g**. The modification of the SuperWavelet is made by the minimization of the energy function $E = \min \|\Psi - g\|_2$ using the Levenberg Marquardt method

Step 2: Generalize these changes on all the wavelets of the associated network. We obtain networks with new wavelets $(\Psi_i^{'}, W_i)$

Step 3: The connections weights of every network are recomputed by the method of the projection of **g** on the dual wavelets to have a new network $(\Psi_i^{'}, W_i^{'})$.

Step 4: Every network $(\Psi_i^{'}, W_i^{'})$ is compared to its origin (Ψ_i, W_i) by computing the Euclidian distances.

Step 5: We compare then these distances. Generally, the smallest one concerns the searched person.

5. Experimental Results

In this section we will 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, we compared it to other methods of face recognition (neural networks (RBFNN), PCA, LDA, EBGM).

5.1. Experiments on ORL face database

The ORL face database contains 40 peoples with 10 images per person. The images have been taken in different times, with slightly varying lightings, facial expressions (open/closed eyes, smiling/non-smiling) and with or without accessories (glasses for example) [7].

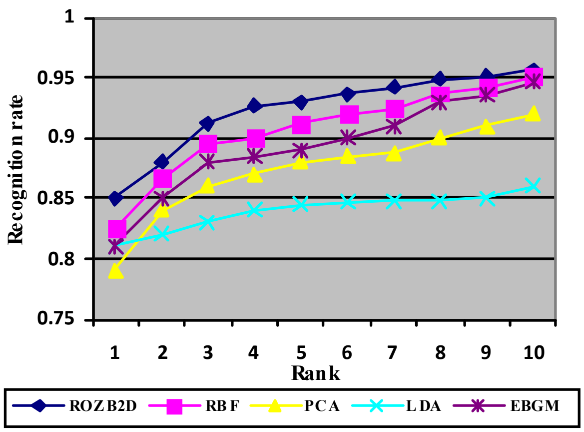

Figure 9. Recognition rate of the ORL database

5.2. Experiments on FERET face database

For more studies, we used the standard FERET data set including the data partitions (subsets) for recognition tests. The gallery consists of 1196 images and there are four sets of probe images (fb, fc, dup1, and dup2) that are compared to the gallery images in recognition stage. We present here, as example, the results of the fb probe.

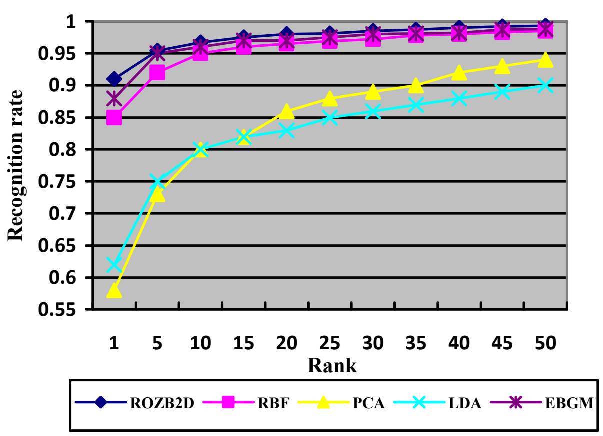

Figure 10. Recognition rate of the (the FERET basis, fb set)

6. Conclusion and future work

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 frames 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 the Superwavelet of the training networks to approximate the face to be recognized.

The performances of the Beta wavelets networks as classifiers 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 the frames theory increased these performances. We are looking actually to optimize it by choosing, at every learning stage, the wavelet from the library that gives the best network output.

The work presented in this article is considered as a stage of a research devoted to the study of digital human face by (3D recognition, face tracing, face detection,…) the Beta wavelet networks, which probably allow us to reinforce the ideas developed at the end of the design of this system.

7. References

[1] I. Daubechies, “Ten Lectures on Wavelets”, *Society of Industrial and Applied Mathematics*, 1992.

[2] Q. Zhang and A. Benveniste. “Wavelet networks”. *IEEE Trans. Neural Networks*, 3:889–898, 1992.

[3] C. Ben Amar, M. Zaied and M.A. Alimi, “Beta Wavelets Synthesis and application to lossless Image Compression”. *Advanced in engineering software Elsiever Edition* 2005

[4] M. Zaied, C. Ben Amar and M.A. Alimi “Award a New Wavelet Based Beta Function”, *International Conference on Signal, System and Design, SSD03*, Tunisia, Mars, 2003, vol. 1, P. 185-191.

[5] H. Szu, B. Telfer, and S. Kadambe. “Neural network adaptive wavelets for signal representation and classification”. *Optical Engineering*, 1992.

[6] V. Kruger and G. Sommer. “Gabor wavelet networks for object representation”. Technical Report CS-TR-4245, University of Maryland, CFAR, May 2001.

[7] H. Tang, M. R Lyu. And I. King, “Face recognition committee machine”, The Chinese University of Hong Kong, 2004.