A Proposed Approach for Image Compression based on Wavelet Transform and Neural Network

Houda Chakib, Brahim Minaoui, Mohamed Fakir, Abderrahim Salhi, Imad Badi
Department of Computer Sciences
Sultan Moulay Slimane University
Beni Mellal, Morocco

Abstract—**Over the last years, wavelet theory has been used with great success in a wide range of applications as signal de-noising and image compression. An ideal image compression system must yield high-quality compressed image with high compression ratio. This paper attempts to find the most useful wavelet function to compress an image among the existing members of wavelet families. Our idea is that a backpropagation neural network is trained to select the suitable wavelet function between the two families: orthogonal (Haar) and biorthogonal (bior4.4), to be used to compress an image efficiently and accurately with an ideal and optimum compression ratio. The simulation results indicated that the proposed technique can achieve good compressed images in terms of peak signal to noise ratio (PSNR) and compression ratio (τ) in comparison with random selection of the mother wavelet.**

Keywords—*Haar wavelet transform; biorthogonal wavelet; backpropagation neural network; scaled conjugate gradient algorithm*

I. INTRODUCTION

Digital image compression is a topical research area in the field of image processing due to its large number of application such as aerial surveillance, reconnaissance, medicine and multimedia communications. Therefore, it has received significant attention of researchers whose major focus is to develop different compression schemes that provide good visual quality with fewer bits to represent digital images in order to secure and facilitate the data transmission by reducing the memory required for its storage [1], [2]. The main core of image compression technology consists of three important processing stages: pixel transforms, vector quantization and entropy coding. Every stage has been considered by the researchers in order to create image compression systems which yield high compression ratio while maintaining high-quality images [3]-[5].

Some of the first papers on wavelet image compression present an excellent performance and support the use of wavelet transform in image compression as it lead to high compression with low distortion [6]-[8]. Moreover, several wavelet families are available for image compression [9] and selecting the appropriate one is very important as many works have proved that the choice of the best wavelet has significant impact on the quality of compression [10]-[12].

Neural networks approaches used for data processing have proved their efficiency due to their structures which offer parallel processing of data and training process makes the network suitable for various kind of data. Furthermore, different image compression techniques were combined with neural network for various applications mainly those based in wavelet transforms [13], [14]. The neural network implementation in every stage of an image compression system was received an important attention from scientists and many works have emerged during the last decade proving that the combination between artificial neural network and wavelet transform is a valuable tool for image processing [15]-[18].

The aim of this present work is to develop an efficient image compression system which combines the features of both wavelet transform and backpropagation neural network in order to find the most suitable wavelet function, out of two functions: orthogonal (Haar) and biorthogonal (Bior4.4), to be used to compress a particular image.

Based on our hypothesis, a neural network can be trained to relate an image contents to their ideal compression function and their optimum compression ratio, just by learning the nonlinear relationship between image pixel values. Once the ideal wavelet function and the highest compression ratio are chosen by the neural network than the image can be compressed and transmitted in an efficiently way.

The remained of the paper is organized as follow: Section 2 presents a brief theory of wavelet transform and the parameters used to evaluate the performance of an image compression system. Section 3 focuses on the theoretical concept of the feedforward neural network. Section 4 briefly presents the backpropagation algorithm used in this work. Section 5 describes in details the proposed method for image compression, including all the components and techniques involved. Section 6 reports experimental results and finally Section 7 provides concluding remarks and possible directions of future research in this area.

II. DISCRETE WAVELET TRANSFORM

For image compression, wavelet transform have recently emerged as a powerful tool that compress images at higher compression ratio. They are a time-frequency representation that takes in account both the time and the frequency of the signal to analyze [2]. The wavelet decomposition of function consists of a sum of functions obtained from simple dilatation and translation; operations done on a main admissible function called "mother wavelet" which must have a compact support and must satisfy the properties of oscillation, translation and

dilatation so other wavelets will be generated [4], [5] as illustrated by the following equation:

$$WT(a, b) = \frac{1}{\sqrt{a}} \int_{t=-\infty}^{t=+\infty} f(t)\psi^*_{a,b}(t)dt \qquad (1)$$

Where,

- $\psi_{a,b}(t) = \Psi\left(\frac{t-b}{a}\right)$ (2)

  - $\Psi(t)$ is the mother wavelet
  - a is the scale parameter
  - b is the translation parameter

As the Fourier transform, wavelet transform is applicable to both continuous as well as discrete signals like digital images. The wavelet transform decomposes the digital image into sub-band in the horizontal and vertical directions. Low and high pass filters are applied to the image along rows and columns separately from this emerge three detailed sub-images: horizontal high-pass sub-image, vertical high-pass sub-image and diagonal high-pass sub-image and one approximate low-pass sub-image [5], [7], as shown in Fig. 1 [9]. The decomposition process is then repeated on the low-pass sub-image to create the next level of the decomposition till the level wanted.

The compression by wavelet consists, in the first set, in image decomposition on a basis of orthogonal or biorthogonal functions, in the second set a quantization technique is applied to the spectral coefficients such as scalar or vector quantization method and finally a binary coding, like Huffman coding, will be applied to convert information under binary shape [1]-[3]. During the coding step the number of pixels to represent the digital image will be reduced consequently the rate compression will be increased.
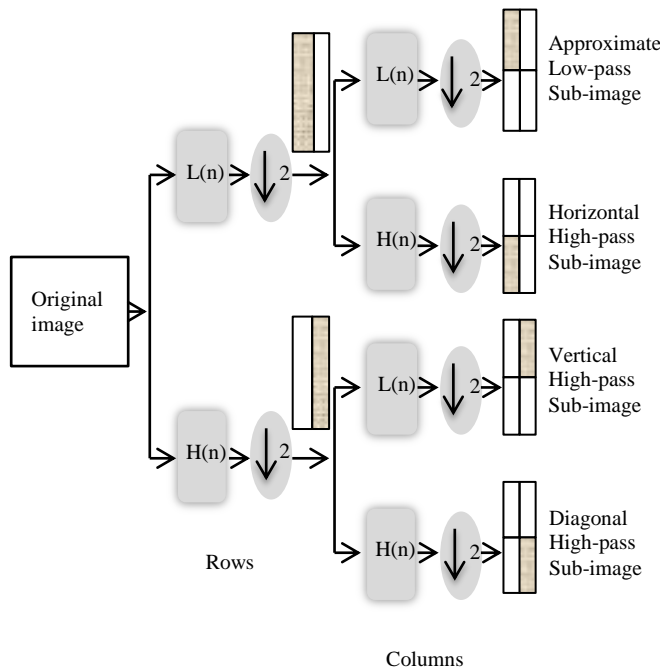

Fig. 1. Wavelet decomposition in one level.

The challenge of an ideal compress system for an image is to find the best compromise between a weak compression ratio and a good perceptual result which determined in terms of the following image quality metrics:

- MSE (Mean Square Error) is the difference between the original and the reconstructed image pixels defined as [16]:

$$MN = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(x_{ij} - y_{ji})^2 \qquad (3)$$

Where, MN is the total number of pixels in the image.

- PSNR (Peak Signal to Noise Ratio) measures the perceptual quality of the image compressed on the term of distortion defined by [7], [16]:

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \qquad (4)$$

- τ is the compression ratio defined as [6], [10]:

$$\tau = \left(1 - \frac{N}{M}\right) \times 100 \qquad (5)$$

Where, - N: number of pixels in compressed image.

M: number of pixels in original image.

- BPP (Bits Per Pixel) gives the number of bits used to store one pixel of the image. For a grayscale image, the initial BPP is 8 bits [1], [3].

III. FEED FORWARD NEURAL NETWORK

The work of an artificial neural network is inspired from the human brain work with a basic building block called neuron or unit [13]. The neuron receives information from an external area and makes a weighted sum of its inputs and calculates by a nonlinear function to generate one output which constitutes one input to another neuron [19]. Every connection between two neurons called weight has different strength [13], [18], [20] and every neuron is connected to other neurons to create the network's layers which implies that neural networks are typically organized in layers and operate according the following scheme: patterns are presented to the network via the input layer which communicates to one or more hidden layers then link to an output layer where the answer is given.

A feed forward neural network is one of the neural network topologies in where data flow through the network layers from the input layer to the output layer in one direction, there is no feedback between layers [13], [20], [21]. This network has fixed inputs and outputs and is used mainly to solve, with high accuracy and generalization ability, problems which are not linearly separable like pattern recognition, classification, prediction and approximation [19], [21]. In other hand, this architecture of neural network learns by examples, it is trained to reproduce in the output the information given in the input, and during training stage, the weights are iteratively adjusted to minimize the distortion between the input vectors and the output vectors for a given error criteria [17], [22].

IV. BACKPROPAGATION ALGORITHM

In this work we used a supervised backpropagation classifier (BPNN) based on the scaled conjugate gradient algorithm because of its implementation simple, the availability of sufficient database for training and its memory requirement is relatively small [23]. On the other hand, the scaled conjugate gradient algorithm fared better in terms of time execution than the other backpropagation algorithms [24].

The backpropagation algorithm (BPNN) is probably the most popular technique used during network training. The basic idea of (BPNN) is to find the percentage of contribution of each connection weight between two neurons to the error in order to correct it by a learning rule [12], [21]. The error propagation consists of two passes through the different layers of the network, a forward pass and a backward pass. In the forward pass the input vector is applied to the network nodes and its effects flows through it layer by layer finally a set of random output is produced as the actual response of the training network and subtracted from the desired response to produce an error, according to it the network makes an appropriate adjustment to its connection weights. This error is measured by the quadratic cost function defined by the following equation [21], [22]:

$$E = \frac{1}{2} \sum_{i=1}^{N} \left(y_{di}(t) - y_i(t) \right)^2 \tag{6}$$

Where,

$y_i(t)$: is the output given by the network.

$y_{di}(t)$: is the desired output called the target.

During the backward pass, the synaptic weights of the network are all adjusted in accordance with an error correction rule to optimize the network performance and to minimize the mean quadratic cost function named also loss function, defined by the equation below [21], [22]:

$$L = \frac{1}{N} E = \frac{1}{2N} \sum_{i=1}^{N} \left(y_{di}(t) - y_i(t) \right)^2 \tag{7}$$

Where, N is the number of patterns presented to the neural network.

As it was mentioned previously, the BPNN implemented for this work, is based on the use of the scaled conjugate gradient algorithm. Ordinary, this algorithm is less fast but more efficient for training large networks [24], [25] and it performs better on function pattern recognition which is our case. The basic idea of the scaled conjugate gradient algorithm is that during training the network the updating of the weight parameters is performed in the direction in which the performance function of the network decreases most rapidly: the negative of gradient and by calculating the gradient function of the quadratic cost function L with respect of the weights. The partial derivation of the loss function with respect to a weight w is given by [21], [26]:

$$\frac{\partial L}{\partial w} = \frac{1}{N} \frac{\partial E}{\partial w} = \frac{1}{2N} \frac{\partial}{\partial w} \left[\sum_{i=1}^{N} \left(y_{di}(t) - y_i(t) \right)^2 \right] \tag{8}$$

Moreover, the training algorithm used falls to the category of batch mode called also off-line training which means that all the weights values are updated after all the training patterns are presented to the network [21]. This training method may assure the stability of our algorithm than the on-line method as it can avoid the creation of series of drawbacks caused by the change of the weights after each training pattern which can possibly resulting in an early stopping of training before all the patterns are presented to the network [21], [23], [26].

V. PROPOSED METHOD FOR IMAGE COMPRESSION

The idea of the work presented in this paper is to train an artificial neural network to relate a specific grayscale image to its ideal compression technique with an optimum compression ratio. The network receives the image contents determined by learning the nonlinear relationship between the pixels intensities and according to this information given, it decides which wavelet Haar or bior4.4 is the most useful for compressing this image and recognizes the optimum compression ratio CR out of the 9 ratios: 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% and 90%.

A. Image Database Compression System

The first step in this study was to collect 120 images of various objects, contrasts and intensities and to convert them to grey level. Hence the image pixels will constitute the nodes of the input layer of our network it was necessary to resize all images into (32x32) pixels in order to reduce the number of input layer neurons and consequently the training time.

The backpropagation neural network BPNN using for this work is a supervised learning network which involves a teacher who provides it with the answers in the form of the target output matrix. The BPNN comes up with guesses while recognizing and compares it with the teacher's "correct" answers and makes adjustments according to errors. The teacher's answers can be established by calculating for every image, the contrast weighted entropy CE which measures the nonlinear relationship between the intensities of an image also it is connected to the image entropy consequently to the information presents within the image. In other meaning, the CE calculates the change of the contrast values between neighbor pixels and it is defined by the following equation [19]:

$$CE = -\sum_{i}^{N} (p_i - m) \times p_i \times \log_2(p_i) \times R \tag{9}$$

Where,

- R: the number of pixels within the image.
- N: number of the intensity values of the pixels.
- p_i: probability of the intensity values.
- m: mean value of the frequencies of the intensities values.

Indeed, as the wavelet transforms are a loss compression techniques [6], [10] which means that information can be lost during the compression process, an image with high CE should be compressed with a low compression ratio τ because there is a lot of information that can be lost if τ is high whereas an image with low CE can be compressed with high

compression ratio τ. Fig. 2 presents some of original images used for this work and their CE coefficients.

B. Implementation of the Neural Network

After the 120 images has been collected and handled, the next step was to create and configure our neural network as mentioned before. Our network is a multilayers classifier with three layers: one input layer with 1024 neurons representing the image pixels, one hidden layer which has been decided, after several experiments, to contain 60 nodes as it assured the perfect network's performance, and finally one output layer with 18 neurons shared as follow: the first 9 neurons represent the 9 Haar compression ratio and the last 9 neurons represent the 9 Bior4.4 compression ratio. A block diagram of our network is represented in Fig. 3.

To train the neural network, the data were randomly divided into three sets as follow:

- Training image set contained 72 images with known ideal compressed method and optimum compression ratio. These images were presented to the network

during training and the network adjusted its error according to them.

- Validation image set which contained 24 images used to measure network generalization and to halt training when generalization stops improving.

- Testing image set contained 24 images with unknown compressed method. These had no effect on training and provided an independent measure of network performance during and after training.

The input data were repeatedly presented to the network and for every presentation the output of the network was compared to the target in order to calculate an error. This error was used to adjust the weights so that finally the model became closer to the reproduction of the target.

We started with initializing the parameters of the network randomly and during training stage, they were iteratively adjusted. We repeated the training task until to achieve a satisfied network performance that hence good training of the network.
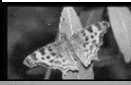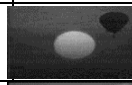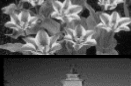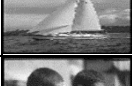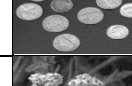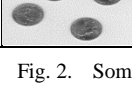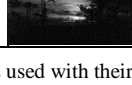
Original image	CE	Original image	CE	Original image	CE	Original image	CE
	-287191.26		-6968686.29		-5957081		-5825.42
	-11045.39		-8024.81		-8591882.81		-21947559.92
	-375062.12		-4228.12		-6944.21		-2150289.31
	-4443.11		-1675197.72		-671150.41		-4519.72
	-1720238.80		-1619143.25		-395664.54		-78030.90
	-7992.19		-5295.83		-7801.90		- 4369.06

Fig. 2. Some original images used with their CE Coefficients.

Fig. 3. Block diagram of the implemented neural network BPNN.

VI. RESULTS AND DISCUSSION

As mentioned before, the neural network used in this work was a backpropagation neural network based on the scaled conjugate gradient algorithm training. This network's topology uses 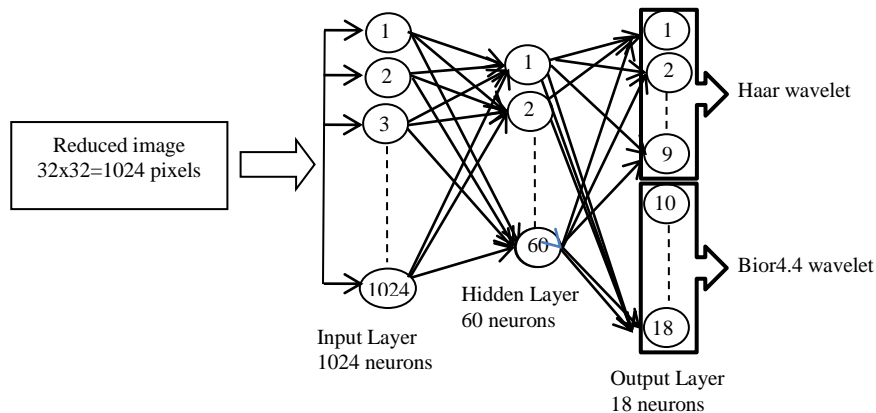sigmoid transfer functions in both the hidden layer and the output layer. During the learning phase, the network parameters were initialized randomly and adjusted to optimize network performance evaluated by the MSE function (mean square error), which is the average squared difference between outputs and targets and the magnitude of the performance function gradient.

After several experiments, the number of hidden layer neurons was set to 60 which assured meaningful training explaining by the low value 6.06×10^{-6} reached by the performance function. Fig. 4 shows the value of performance function during learning versus iteration numbers including training, validation and test curve. As we can see in this figure, the training is acceptable because first the final mean square error is very low and stabilize after 55 iterations (epochs) and the best validation performance is 0.052 reached in the 49th iteration.

In the second hand the validation curve and the test curve converge in the same way.

Moreover, The neural network learnt and converged after 55 epochs within 33s and it recognized correctly all the 72 training images thus yielding 100% recognition of the training set whereas only 10 out of 24 images from the test set with unknown ideal compression method and optimum ratio are correctly classified which yielded only 40% recognition which is not sufficient.

Another most interest network parameter is the magnitude of performance gradient represented in Fig. 5 which shows the gradient function versus iteration numbers during the neural network training. As shown in this figure the gradient became very small and closer to zero: it reached the value 6.98×10^{-5} after 55 iterations also this implies that the performance function was minimized which mean that the outputs were very close to the targets and hence the network was trained with success.

For this work, a minimum accuracy level of 98% was considered acceptable because the network performance function reached the low value of 0.001. With this accuracy, the network learnt correctly 70 out of 72 images from the training set after 71 iterations within 29s. Fig. 6 shows the value of performance function, during training, validating and testing stage, versus iteration numbers.
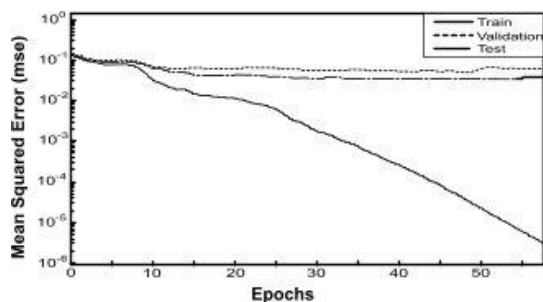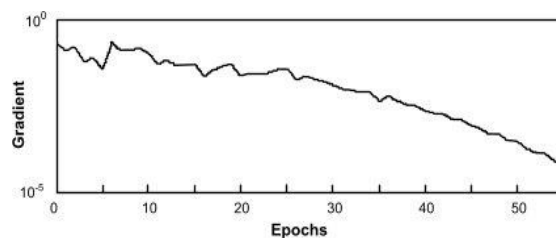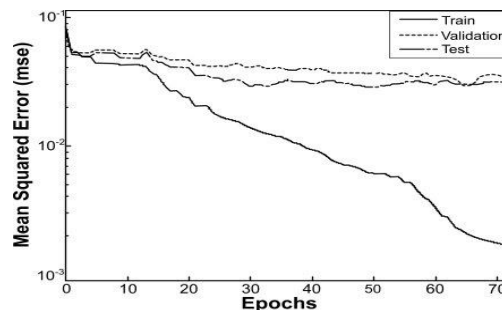

Fig. 5. Gradient curve for accuracy 100%.

Fig. 6. Neural network learning curve for accuracy 98%.

Using this accuracy, the neural network yielded 75% correct recognition rate of ideal compression method and optimum compression ratio for the 24 images from testing set that were not seen before by the network which mean that 18 out of 24 images never presented to the network were correctly classified whereas 6 of them were misclassified.

Furthermore, the gradient of the performance function, represented by Fig. 7, reached the low value 0.001 after 71 iterations which indicate that the outputs given by the network were close to the desired targets and hence the network is trained.

The results of this work are demonstrated in Fig. 8 which shows examples of the optimally images as determined by the trained neural network.

VII. CONCLUSION AND FUTURE WORKS

In this work, we propose to train a neural network to associate a grayscale image to its ideal compression method and optimum compression ratio just by learning the nonlinear relationship between the intensity of the image in order not to lose important information within the original image. The proposed system was developed and implemented using 120 images of various objects, contrasts and intensities.

The neural network used to implement this image compression system learnt to associate all the 72 images from the training set to their ideal compression methods and optimum compression ratio with the accuracy rate of 100%.

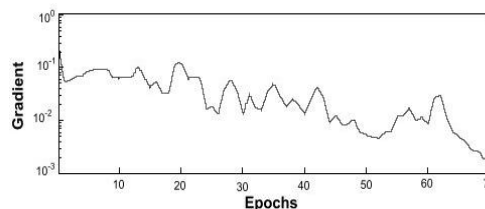Fig. 4. Neural network learning curve for accuracy 100%.

Fig. 7. Gradient curve for accuracy 98%.

Using only the accuracy rate of 98 % the neural network was able to relate 18 out of the 24 images from the testing set to their ideal wavelet with an optimum compression ratio upon presenting the image to the network which represents about 75% recognition rate which is still not very satisfied and not sufficient.

Future work will focus in improving the network performance so that the accuracy rate achieves a value more than 75% in order to implement this compression system for medical images as the compression of this kind of images receives an important attention from researchers in many fields.
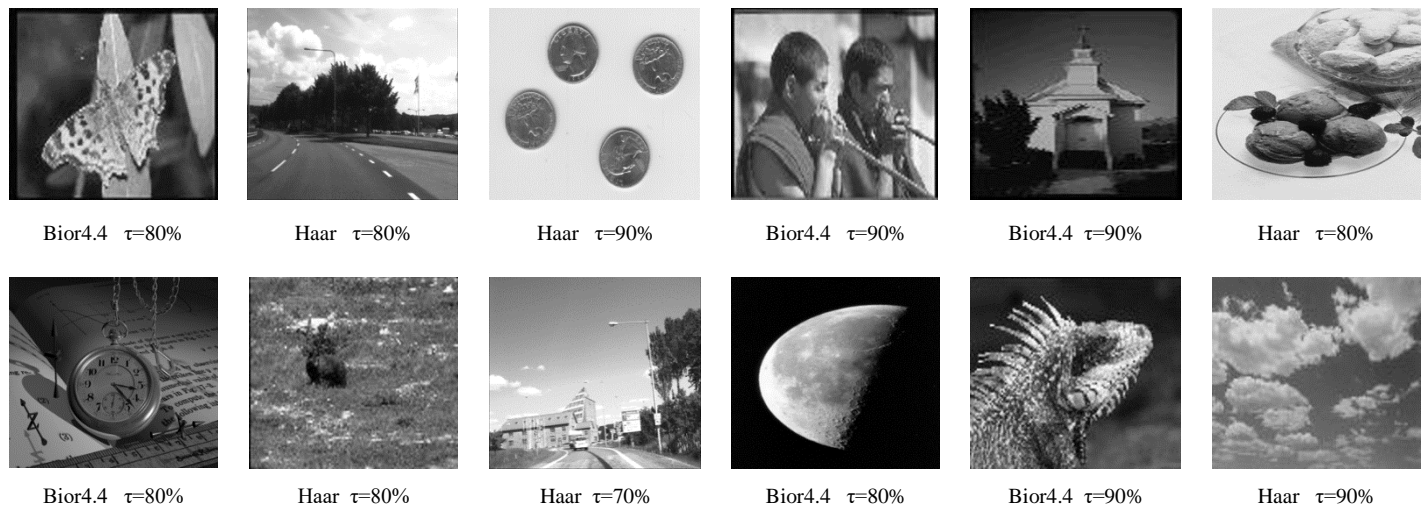
Bior4.4 τ=80%   Haar τ=80%   Haar τ=90%   Bior4.4 τ=90%   Bior4.4 τ=90%   Haar τ=80%

Bior4.4 τ=80%   Haar τ=80%   Haar τ=70%   Bior4.4 τ=80%   Bior4.4 τ=90%   Haar τ=90%

Fig. 8.   Some results of this application.

REFERENCES

[1] M. Rabbani, and P.W. Jones, "Digital Image Compression Techniques," vol.7, SPIE Press, Bellingham, Washington, USA, 1991.

[2] P.C. Cosman, R.M. Gray, and M. Vetterli, "Vector Quantization of Image Sub-bands: A Survey," IEEE Trans. Image Processing, vol. 5, 1996, pp. 202-225.

[3] R.E. Crochière, S.A. Webber, and J.L. Flanagan, "Digital Coding of Speech in Sub-bands," Bell syst, Tech. J. vol. 55, 1976, pp. 1069-1085.

[4] S.G. Mallat, "Multifrequency Channel Decomposition of Images and Wavelet Models," IEEE trans. Acoust, speech, Signal processing, vol. 37, 1989, pp. 2091-2110.

[5] J.W. Woods, and S.D. O'Neil, "Sub-bands Coding of Images," IEEE Trans. Acoust, Speech, Signal Processing, vol. 34, 1986, pp. 1278-1288.

[6] S.G. Mallat, "A theory for Multi Resolution Signal Decomposition: The wavelet Representation," IEEE Trans. Pattern Anal. Machine Intell, vol. 11, 1989, pp. 674-693.

[7] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding using Wavelet Transform," IEEE Trans. Image Process. 1 (2), 1992, pp. 205-220.

[8] A. Said, and W.A. Pearlman, "A new Fast and Efficient Image Codec based on Set Partitioning in Hierarchical Trees," IEEE, Trans. Circuits syst. Video Technol. 6 (3), 1996, pp. 243-250.

[9] T. Gandhi, B.K. Panigrahi, and S. Anand, "A Comparative Study of Wavelet Families for EEG Signal Classification," Neurocomputing 74, 2011, pp. 3051-3057.

[10] K. Ratakondu, and N. Ahuja, "Loless Image Compression with Multiscale Segmentation," IEEE, Transactions on Image Processing, vol. 11, N°. 11,2002, pp. 1228-1237.

[11] K. Ahmadi, A.Y. Javaid, and E. Salari, "An efficient Compression Scheme based on Adaptive Thresholding in Wavelet Domain using Particle Swarm Optimization," signal processing: image communication 32, 2015, pp.33-39.

[12] Q. Zang, and A. Beneveniste, "Wavelet networks," IEEE Tans. Neural Networks, vol. 3, 1992, pp. 889-898.

[13] S. Osowski, R. Waszczuk, and P. Bojarczak, "Image Compression Using Feed Forward Neural Networks- Hierarchical Approach," Lecture Notes in Computer Science, Book chapter, Springer – Verlag, vol. 3497, 2006, pp. 1009-1015.

[14] Q. Zang, "Wavelet Network in Nonparametric Estimation," IEEE Trans. Neural Networks, 8(2), 1997, pp. 227-236.

[15] A.V. Singh, and K.S. Murthy, "Neuro-Wavelet based Efficient Image Compression Using Vector Quantization," International Journal of Computer Applications (0975-08887), vol. 49-N°.3, July 2012.

[16] T. Denk, K. Perhi, and V. Cherkassky, "Combining Neural Network and the Wavelet Transform for Image Compression," Proceeding of Intl Conf, 1993, pp. 637-640.

[17] V. Krishnanaik, G.M. Someswar, K. Purushotham, A. Rajaiah, "Implementation of Wavelet Transform, DPCM and Neural Network for Image Compression," International Journal of Engineering and Computer science ISSN: 2319-7242, vol. 2, issue. 8, August 2013, pp. 2468-2475.

[18] J. Jiang, "Image Compressing with Neural Networks – A Survey," signal processing: image communication, vol. 14, N°9, pp. 737-760, 1999.

[19] K. Dimililer, "Backpropagation Neural Network Implementation for Medical Image Compression," Journal of applied mathimatics, Vol. 2013, Article ID 453098, November 2013, pp. 1-8.

[20] K. Dimililer, A. Khashman, " Image Compression using Neural Networks and Haar wavelet," Transaction on Signal Processing, ISSN: 1790-5052, vol. 4, issue: 5, May 2008.

[21] A.K. Alexandridis, A.D. Zaprani, "Wavelet Neural Networks: A Pratical Guide," neural networks 42, 2013, pp. 1-27.

[22] C. Ben Amar, O. Jemai, "Wavelet Networks Approch for Image Compression," ICGST, GVIP special issue on image compression, 2007.

[23] R.D. Dony, S. Haykin, "Neural Network Approaches to Image Compression," IEEE, Proceedings, vol. 83, N°2, pp. 288-303, Février 1995.

[24] D. Batra, "Comparison between Levenberg-Marquardt and Scaled Conjugate Gradient Training Algorithms for Image Compression using MLP," International Journal of Image Processing, vol. 8: issue (6), 2014.

[25] H. Szu, B. Telfer, and J. Garcia. "Wavelet Transforms and Neural Networks for Compression and Recognition," Neural Networks, 9:695-708, 1996.

[26] A.J. Hussain, D. Al-Jumeily, N. Radi, P. Lisboa,"Hybrid Neural Network Predictive-Wavelet Image Compression System," neurocomputing 151, 2015, pp. 975-984.