

Wavelet-based Image Modelling for Compression Using Hidden Markov Model

Muhammad Usman Riaz

MCS, National University of Science
and Technology Islamabad, Pakistan

Imran Touqir

MCS, National University of Science
and Technology Islamabad, Pakistan

Maham Haider

MCS, National University of Science
and Technology Islamabad, Pakistan

Abstract—Statistical signal modeling using hidden Markov model is one of the techniques used for image compression. Wavelet based statistical signal models are impractical for most of the real time processing because they usually represent the wavelet coefficients as jointly Gaussian or independent to each other. In this paper, we build up an algorithm that succinctly characterizes the interdependencies of wavelet coefficients and their Non-Gaussian behavior especially for image compression. This is done by extracting the combine feature of hidden Markov model and Wavelet transformation that gives us comparatively better results. To estimate the parameter of wavelet based Hidden Markov model, an efficient expectation maximization algorithm is developed.

Keywords—Hidden Markov model; Wavelet transformation; Compression; Expectation Maximization

I. INTRODUCTION

Wavelet transformation is the tool for statistical signal processing and image modeling often used in real time signals processing[1]. Due to strong coordination between wavelet coefficients, these models have complicated processing but performance is much better. Wavelet transformation has primary and secondary properties that are used for statistical signal processing and image modeling[2]. These wavelet based hidden Markov models have many real-time applications in Engineering and Medical field.

In statistical signal processing techniques, wavelet transformation consider wavelet coefficient as single scalar coefficient that gives powerful tools for image modeling, previously discussed in [3,4,5]. These techniques consider coefficients as independent to each other but such methods those exploit dependencies between coefficients give better results.

In this paper we have developed a wavelet based hidden Markov model that succinctly model the statistical dependencies of the wavelet coefficients that are non-Gaussian in nature. Proposed algorithm exploits the statistical dependencies of the wavelet coefficients for better compression results. For this purpose we have used the combine properties of the wavelet transformation and Hidden Markov model.

A. Wavelet Transformation

Wavelet transformation is used to convert signals or images into its coefficients that contains complete information about

signal or image [6]. Wavelet transformation converts the image into four subparts, first part contains approximation of the original image and remaining three parts contain the diagonal, horizontal and vertical coefficients information respectively [7].

Complete modeling of the image can be done with the help of wavelet coefficient and scaling coefficients. Primary properties (locality, multi-resolution and compression) of the wavelet transformation plays an important role in approximation of many real time signals [8]. With the help of locality each wavelet atom can be localized in time and frequency domain simultaneously. To analyze the wavelet atom at any scale we can use multi-resolution property of wavelet transformation [9].

Both locality and multi-resolution properties of wavelets facilitate us to match the large range of real time signals [10]. Most of the complicated signals can be approximated with small number of wavelet basis and scaling coefficients which will be discussed thoroughly in coming sections. We can conclude that all those statistical signal processing and image modeling methods that uses wavelet transformation are more beneficent as compared to those methods that uses only frequency-domain or time-domain information of the image/signal.

B. Statistical modelling for image

To approximate the wide range of coefficients we can use probability model based on wavelet transformation that is flexible, rich and tractable[11].

Previously wavelet coefficients modeled as non-Gaussian or jointly Gaussian[12,13] but statistically independent to each other [14]. To capture the linear dependencies of the wavelet coefficients, jointly Gaussian model is used. Histogram of wavelet coefficients density is peaky at zero indexes and heavily tailed than the histogram of the typical Gaussian distributions[15]. Complete decorrelation of wavelet coefficients is impossible, residual dependencies are always present between the wavelet coefficients after wavelet transformation[16]. Non Gaussian Models that doesn't exploit the complete statistical dependencies of wavelet coefficients during modeling are unrealistic to work with.

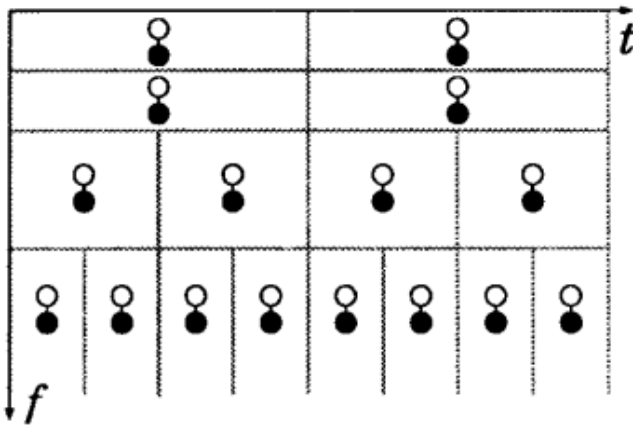


Fig. 1. Independent Mixture Model, each black and white node represent the continuous wavelet coefficients C_i and mixture state variable Q_i for C_i respectively

To overcome these problems, secondary properties of the wavelet transformation is helpful[17,18]. Clustering(one of the wavelet transformation property) suggests that for any high/low energy wavelet coefficient, neighboring coefficient will also be high/low energy coefficients[19]. According to persistence, values of the wavelet coefficients propagate across the scale. We will use these properties in the statistical modeling of the image.

To completely characterize the statistical dependencies of all wavelet coefficients we need to model the jointly probability density function that would consider all the dependencies of the wavelet coefficients but it is intractable and impossible to apply on real time images[20]. Conversely, modeling the wavelet coefficients without exploiting dependencies of the wavelet coefficients is simple and easy to implement but would not consider the inter-coefficients dependencies[21]. For better performance we need to make a balance between these two extremes.

We developed a wavelet based hidden Markov model that completely characterizes the probability structure of the wavelet coefficients. We model the marginal probability of each coefficient as mixture density with state variable to match the non-Gaussian nature of the wavelet coefficients as shown in figure 1. We introduce the Markovian dependencies between the hidden state variables to characterize the key dependencies between the wavelet coefficients. These dependencies are illustrated in figure 1.

Hidden Markov model is one of the Probabilistic graphical model used in statistical signal and image modelling. In this paper we used three probabilistic graph models with state-to-state connectivity as shown in figure 2. Independent mixture model ignores statistical dependence of wavelet coefficient and leave the variable states unconnected. Within each scale, to connect the state variable horizontally, we used the hidden Markov chain model. To connect the state variable vertically across the scale hidden Markov tree model is used. Together these three models represent the wavelet based hidden Markov Model.

Remaining distribution of the paper is as: Section 2 discusses the definitions and notation that will be used in mathematical modeling. Section 3 explains the statistical image modeling using hidden Markov model. Section 4 will elaborate the EM Algorithm and training of the proposed algorithm. Application of proposed framework will be discussed in section 5 along with experimental results. At the end section 6 will conclude the complete paper and suggest the future work.

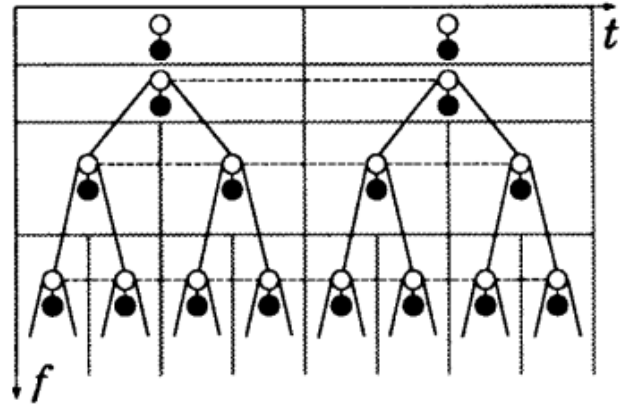


Fig. 2. Horizontal dashed line represent the connection capitate HMC model and vertical solid line represent the connection capitate HMT model respectively

II. NOTATIONS

Wavelet based hidden Markov model can completely characterize with the help of trees and graphs. $\{n_1, n_2, n_3, n_4 \dots \dots n_N\}$ used to represent the graph nodes. Term “connection” used to link the nodes. Ancestors are those nodes that are present on the path going from v_i to the root node. Descendants are those nodes that are present in the path from v_i going away from root node. $n_p(i)$ is parent of n_i node if it is immediate ancestor. To denote the children of n_i we will represent it as $\{n_j\} j \in c(i)$. A node has many children but have only one parent. In case of binary tree one parent has only two immediate children. We will represent the c_i^k , i^{th} wavelet coefficient from the k^{th} tree. In this paper capital letters will be used for random variables (R.V) and small letters will be used as an observed value of that R.V. Probability mass function of discrete random variable “Q” denoted as $p_Q(q)$ and $f_C(c)$ as probability density function of continuous R.V “C”.

III. WAVELET-DOMAIN PROBABILITY MODELS

Main problem is to exploit the key dependence of the wavelet coefficients during modeling of wavelet based hidden Markov model by considering the wavelet coefficients that follows non- Gaussian distribution. This modeling is done in two steps. Initiate with simple statistical model with the assumption that all wavelet coefficients are independent to each other based on the fact that wavelet transform de-correlate many of the real time signal’s coefficients. Then enlarge this model using Markovian structure to exploit the residual dependencies of the wavelet coefficients. Here hidden Markov model is helpful that uses the state of the wavelet coefficients instead of values of the wavelet coefficients. Both first order Markovian dependencies and marginal Gaussian mixture provides to the implementation this model practically.

A. Wavelet Transformation

Primary properties of the wavelet transformation allow us to model each wavelet coefficient as high or low state. High state represents all those components carrying significant amount of information and low state is for less energy wavelet components. So, for each wavelet coefficient we have two states. 0-mean density and high variance is for high state whereas, 0-mean density and low variance for low state wavelet coefficients. Each wavelet coefficient is modeled as a 2-state mixture model illustrated in fig 3.

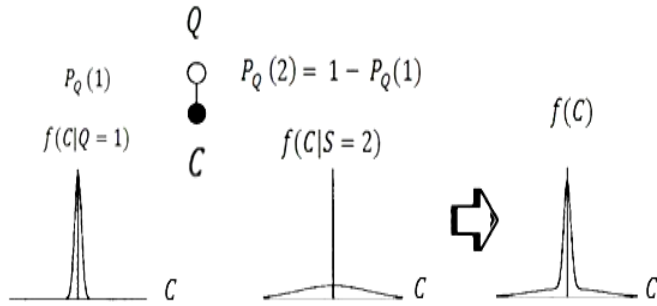


Fig. 3. Two-state, 0-mean GMM. White node represent the state variable and black node is for random variable. $p_Q(1)$ Represent low variance (state 0) and $p_Q(2)$ represent high variance (state 1)

Now we can completely characterize the 2-state, 0-mean with probability mass function of $p_Q(1)$ and $p_Q(2)$. Mostly Q (state variable) is hidden while C (value of coefficient) is observed.

For simplicity we will focus only on 2-state model. Consistent with $p_Q(q)$ (probability mass function) “ Q ” can have only two values.

The probability density function of W is:

$$f_C(c) = \sum_{m=1}^M p_Q(m) f_{c|Q}(c|Q=m) \quad (1)$$

However, as “ Q ” is a random state variable and its value is hidden to us so probability density function will be non-Gaussian and overall wavelet coefficient has non-Gaussian density function even though wavelet coefficients are conditionally Gaussian. Normally scaling coefficients has non-zero variance so it is inappropriate to use 2-state, 0-mean Gaussian mixture model. Midway is to use Gaussian mixture model but with mixing densities that have non-zero means.

Wavelet transformation de-correlates the wavelet coefficients of many real time signals. So for statistical modeling of wavelet coefficients, independent Gaussian mixture model gives considerable improvement over deterministic signal. We have modeled one wavelet coefficient using 2-state Gaussian mixture model so it looks logical to use it for all the wavelet coefficients. Preferably, we need such a probabilistic model that collectively consider the probability density function of wavelet coefficients and exploit the statistical dependencies of the wavelet coefficients. After developing the Gaussian model for one coefficient, we extend the Gaussian mixture model to two coefficients with the help of jointly Gaussian mixture model.

Persistence and clustering recommended that if any wavelet coefficient is in state one (high state) then most probably its neighboring wavelet coefficient will also be in high state. Two such wavelet coefficients that are in neighbor can be modeled using Gaussian Mixture model variables that are in independent state. This simple modeling ensures the modeling of overall wavelet coefficients using the same method.

B. Wavelet Transformation Based Graph Models

To make a connection between states and wavelet coefficients, primary properties of wavelet transform are helpful. To represent the dependencies of wavelet coefficients there is link connected between variables. We have three ways to connect the dots and will be discussed later in this section. To make horizontal and vertical connection between Q_i “state variable”, HMT model and HMC model are used respectively. Transition probability gives information about the probability of transition from one state to the other. We can model dependencies of state variable using Hidden Markov model.

Parameter of hidden Markov model is defined as:

- $p_{Q_i}(m)$, Probability mass function for Q_i (root node).
- $\epsilon_{i,p(i)}^{mr} = p_{Q_i|Q_{p(i)}}[m|Q_{p(i)} = e]$, conditional probability of Q_i given $Q_{p(i)}$ is in the state of e .
- $\mu_{i,m}$ & $\sigma_{i,m}^2$ represent the variance and mean.

Above parameters are collectively called the parameter of the model and represented by “ θ ”. Remember that for this paper we assume that we have two states with 0-mean.

$$f_{C_i}(c_i|\{C_j\}_{j \neq i}, \{Q_i = q_j\}_{j \neq i}, Q_i = q_j = f_{C_i}(c_i|Q_i = q_i) \quad (2)$$

As state of wavelet coefficients are hidden to us so wavelet based Hidden Markov model does not rely only on the wavelet coefficients of Markov structure. Suppose $J(i)$ represents the scale of wavelet coefficient (C_i and Q_i).

$$f_{C_i}(c_i|\{C_j\}_{j(i) > J(i)}) \neq f_{C_i}(c_i|C_{(i)}) \quad (3)$$

Generally, wavelet coefficients are not Markov. Wavelet state variables owe Markov nature that is why wavelet based HMM is efficient for modeling the wavelet coefficients.

IV. EXPECTATION MAXIMIZATION ALGORITHM

To estimate the model parameter “ θ ” of the wavelet based HMM, we need to train the data that consists of wavelet coefficients “ C ” of the image. These parameters include probabilities of mixture state, mean “ $\mu_{i,m}$ ” and variance “ $\sigma_{i,m}^2$ ” of the GMM. We then apply the maximum likelihood principle to find the parameters of the proposed model. Direct estimation using likelihood principle is hard to estimate. Because meaning of the estimation is to find the hidden state “ Q ” of the wavelet coefficients “ C ”, means and variance of the Gaussian Mixture model. Expectation maximization is an iterative algorithm that is used to collectively find the model parameters “ θ ” and probabilities for the unobserved states “ Q ”. Expectation maximization is also known as Baum Welch algorithm in the domain of hidden Markov model. Details of expectation maximization steps in the case of Hidden Markov chain and independent mixture model will be discuss in coming section.

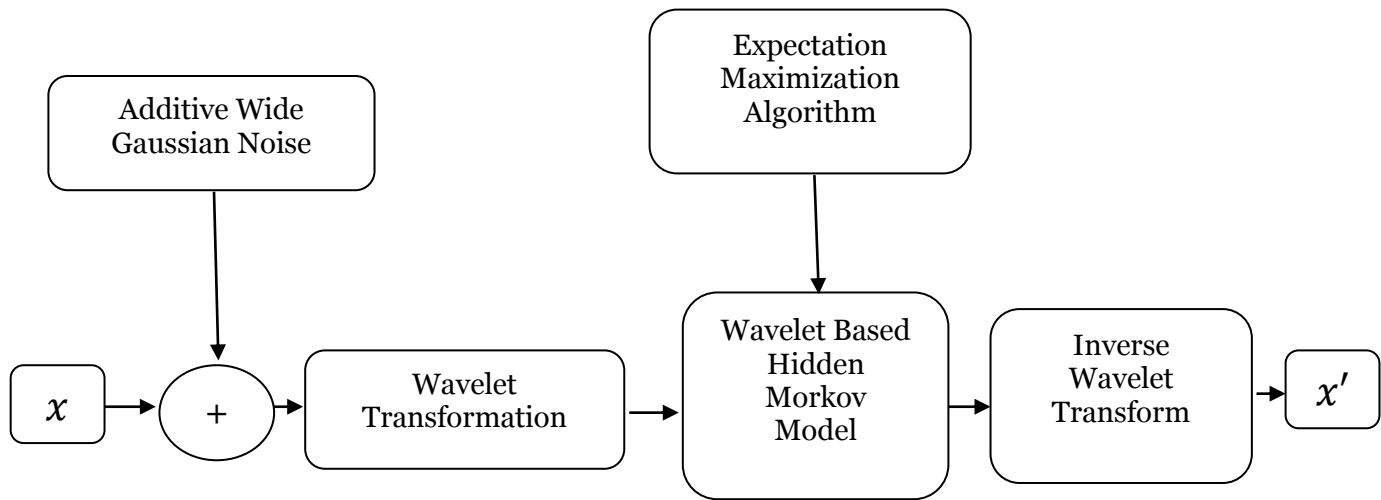


Fig. 4. Block diagram of the proposed wavelet based Hidden Markov Model

Figure 4 illustrate the complete functionality of algorithm graphically that is used to estimate the model parameters of the wavelet based hidden Markov model.

A. Expectation Maximization algorithm for training

In EM algorithm, “Training” is used for fitting the model parameters to the Wavelet based hidden Markov model. Basically training is used to avoid the “over-fitting”. Our objective is to maximize the $\ln f(c|q)$ log function of wavelet coefficients “c” given hidden state variable “q” for this purpose expectation maximization algorithm performs two steps. In expectation step (called E-Step) algorithm, find the $E_S[\ln f(c, Q|\theta)|c, \theta^i]$ and in next step algorithm, maximize what it found in E-Step. Convergence will direct the algorithm in right direction if initialization is correct. Steps of the algorithm are given below:

Selection of a model parameter θ^0

Counter set to be zero, $i = 0$ (initialization)

- E step (expectation step)

Find $p(S|w, \theta^i)$, used in $E_S[\ln f(c, Q|\theta)|c, \theta^i]$ maximization.

- M step (maximization step)

Set $\theta^{i+1} = \arg \max_{\theta} E_Q[\ln f(c, Q|\theta)|c, \theta^i]$

- update $i = i + 1$.

If it converges, then stop the iteration

Otherwise start again from step 1.

Initialization is important in EM algorithm. If we intelligently initialize the algorithm the complexity will reduce and it will start converging in few steps but in case if we randomly initialize, then it might be possible that after few iterations the complexity of algorithm exponentially increase and it will start diverging. To reduce the complexity of the algorithm we need to initialize the algorithm intelligently.

V. EXPERIMENTAL RESULTS

Although this model works for statistical image processing in numerous real time signals. In this paper we have developed a wavelet based statistical model for image compression. We have compared the compression performance of our model by using different types of wavelets. Proposed model has a substantial improvement over previous models. Our results demonstrate the performance of the wavelet based Hidden Markov model in image compression.

TABLE I. PERCENTAGE OF COMPRESSION RATIO AND PSNR IMAGE NAME: JELLY FISH, IMAGE SIZE 128X128, LEVEL 1

S.No.	Wavelet (Level 1)	PSNR (db)	Compression Ratio
1	Bior 3.1	52.85	70.59%
2	Bior 3.3	53.70	70.14%
3	Bior 3.5	53.92	70.06%
4	Bior 3.7	53.98	70.08%
5	Bior 3.9	54.00	70.08%
6	Bior 4.4	56.82	62.28%
7	Bior 5.5	58.07	69.56%
8	Bior 6.8	56.69	72.66%

TABLE II. PERCENTAGE OF COMPRESSION RATIO AND MEAN SQUARE ERRORS IMAGE NAME WOOD STATUE, SIZE 256*256, LEVEL

S.No.	Wavelet type	MSE (db)	Compression Ratio
1	Haar	0.1201	65.85%
2	Sym2	0.1368	63.36%
3	Sym 3	0.1364	61.83%
4	Sym 4	0.1382	61.04%
5	Sym 6	0.1355	60.54%
6	Sym 8	0.1343	60.36%
7	Db1	0.1201	65.85%
8	Db 2	0.1368	63.36%
9	Db 3	0.1364	61.83%
10	Db 4	0.1306	59.31%
11	Db 6	0.1339	61.59%
12	Db 9	0.1340	62.20%
13	Coif1	0.1355	62.01%
14	Coif 2	0.1315	60.30%
15	Coif 3	0.1292	59.60%
16	Coif 4	0.1318	59.96%
17	Coif 5	0.1296	59.63%
18	Dmey1	0.1310	61.06%
19	Dmey 2	0.1110	68.52%
20	Rbio1	0.1123	73.96%
21	Rrbio 2	0.08184	77.30%
22	Bior1	0.1201	65.86%
23	Bior 2	0.1315	66.75%
24	Bior 3	0.1379	66.20%
25	Bior 4	0.1644	58.46%
26	Bior 5	0.1379	58.38%
27	Bior 6	0.1289	59.83%
28	Bior 7	0.1403	56.96%

A. Objective analysis for image (2d) compression using wavelet domain hmm

To achieve the image compression, we apply the different wavelets on proposed algorithm. We had applied 28 different wavelets and observe behavior of proposed algorithm. From Table 2, we can analyze that as the wavelet type changes, the compression ratio and mean square value also changes. Results illustrate that compression ratio depends of proposed algorithm depends on image characteristics as well as on the wavelet type used. When we apply dmey1 wavelet on the wood statue image, size 256x256 it gives 61.06% compression with 0.1310db mean square error (serial number 18 in table 2) and by changing the wavelet types we found different compression ratio as well as mean square error as shown in Table 2. "rbio 2" is the best wavelet for this test image because it provides 77.30% compression ratio with 0.08184db mean square error (serial number 21 in table 2).

Table 3 represent the comparison between different wavelet of same type "Brior" by comparing their means square error and compression ratio. We have found "Bior 6.8" best among all that gives 0.1392db mean square error and 72.66% compression ratio and "Bior 3.1" worst that give 0.337db mean square error and 70.59% compression ratio.

TABLE III. PERCENTAGE OF COMPRESSION RATIO AND MEAN SQUARE ERRORS IMAGE NAME: JELLY FISH, IMAGE SIZE 128X128, LEVEL 1

S.No.	Wavelet type	MSE (db)	Compression Ratio
1	Bior 3.1	0.3377	70.59%
2	Bior 3.3	0.2771	70.14%
3	Bior 3.5	0.2638	70.06%
4	Bior 3.7	0.2603	70.08%
5	Bior 3.9	0.2591	70.08%
6	Bior 4.4	0.1353	62.28%
7	Bior 5.5	0.1013	69.56%
8	Bior 6.8	0.1392	72.66%

B. Subjective analysis for image (2d) compression using wavelet domain hmm

The proposed algorithm is implemented and tested over the wide range of grayscale and colored images. The natural test images used are wood horse, Persons, Wood Statue, Mask, Facets, Laure, Catherine, Wood Statue, Arms and Jelly Fish. The results for these images are given in figure 5.

The goal of subjective image comparisons is to determine the effects of the following on compression performance: compression by using different types of wavelets on grayscale and colored images.

Jelly fish, one of the most common test images in compression research, consists primarily of low frequency content. Table 3 lists the MSE results and Compression ratios using different types of wavelets for jelly fish. The "rbio 2" wavelet depicts the best MSE performance among all.

VI. CONCLUSION AND FUTURE WORK

Wavelet based hidden Markov model is one of the model that is used for image modeling. These models are helpful in statistical modeling of wavelet coefficients of images that succinctly models the coefficients that don't follow the Gaussian distributions.

We have developed the wavelet based hidden markov model for statistical image modeling of the wavelet coefficients for compression. This model allows us to exploit the interdependencies of the wavelet coefficients and consider the entire coefficients during modeling that does not follow the Gaussian distribution. Mostly statistical methods model the wavelet coefficients as jointly Gaussian or independent to each other so, these models exploit less information about image characteristics. For parameter estimation of wavelet based hidden markov model, an efficient expectation maximization algorithm is developed.

Proposed approach gives the encouraging compression results in image compression domain, further research should be directed to multidimensional wavelet domain Hidden Markov Models as in this paper only statistical image modeling for 2-states is discussed.

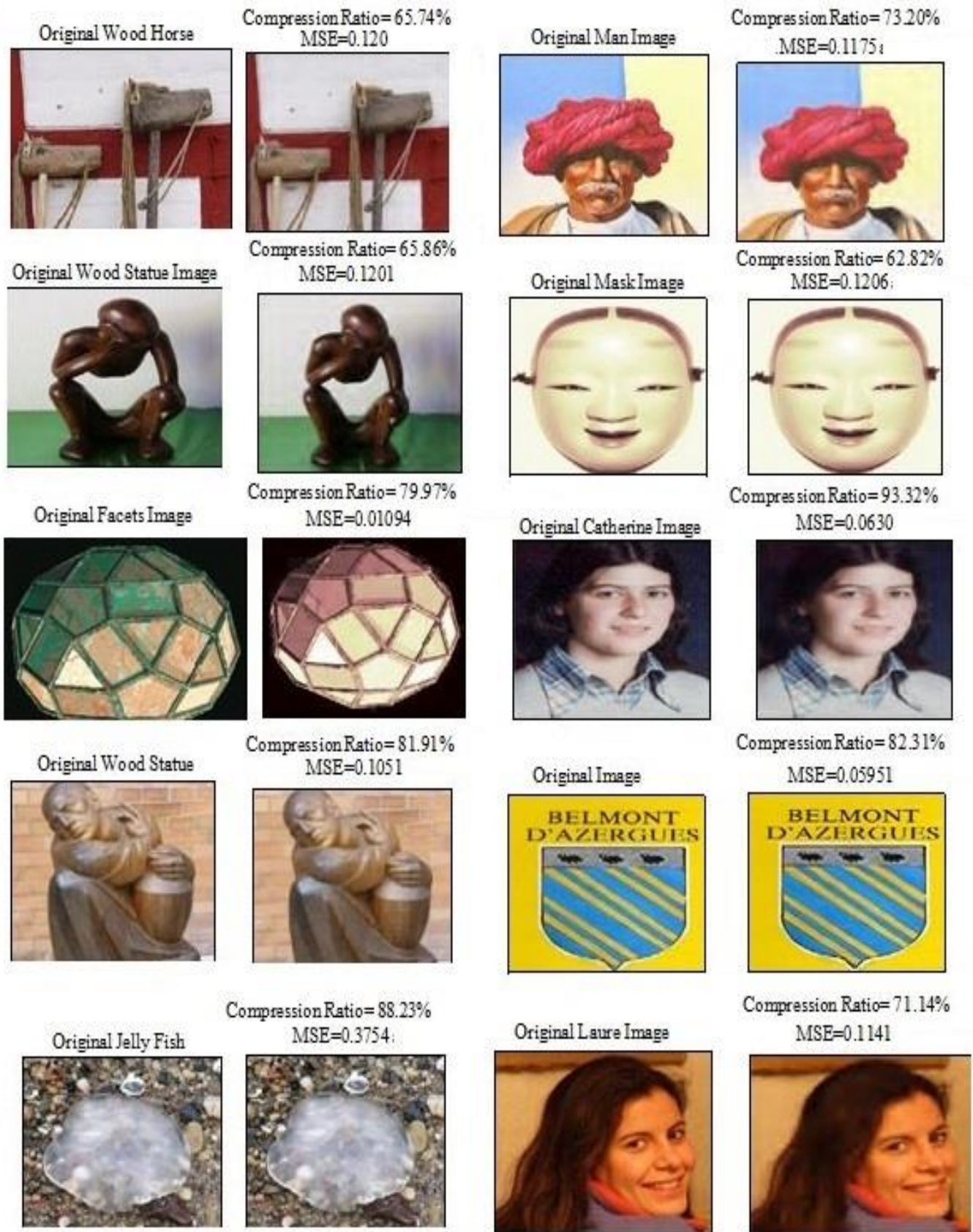


Fig. 5. Comparison between compression ratio and mean square error of different images

REFERENCES

- [1] Camps-Valls, G., Tuija, D., Bruzzone, L., and Atli Benediktsson, J. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Proc. Mag.* 31, 1 (Jan. 2014), 45, 54.
- [2] M. A. Losada, G. Tohumoglu, D. Fraile, and A. Artes, "Multi-iteration wavelet zerotree coding for image compression," *Sci. Signal Process.* vol. 80, pp. 1281–1287, 2000.
- [3] M. S. Crouse and R. G. Baraniuk, "Contextual hidden Markov models for wavelet-domain signal processing," in *Proc. 31st Asilomar Conf. Sig nals, Syst., Comput., Nov.* 1997.
- [4] H. Choi and R. Baraniuk, "Image segmentation using wavelet-domain classification," in *Proc. SPIE*, vol. 3816, Denver, CO, July 1999.
- [5] Gilbert Strang and Truong Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1997.
- [6] Y. U. Khan and J. Gotman, "Wavelet based automatic seizure detection in intracerebral electroencephalogram," *Clin. Neurophysiol.* vol.114, pp. 898–908, 2003
- [7] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. New York: Academic, 1990.
- [8] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.
- [9] Kentaro Kinebuchi, signal interpolation using Wavelet based hidden Markov trees, 2000
- [10] Yaonan Wang, Xiaoping Ma, Application of MultiScale Hidden Markov modeling wavelet coefficients to fMRI activation detection. (Report): An article from: *Journal of Mathematics and Statistics*.
- [11] M. Alam, C.A. Rahman, W. Badawy, G. Jullien, Efficient Distributed Arithmetic Based DWT Architecture for Multimedia Applications, *Proceedings of the 3rd IEEE International Workshop on System-on-Chip for Real-Time Applications*, pages 333 -336, June 2003.
- [12] M.Nibouche, A.Bouridane and O.Nibouche, A Framework for A Wavelet-Based High Level Environment, the 8th IEEE International Conference on Electronics, Circuits and Systems (ICECS), pages 429 - 432, vol.1, Sept. 2001.
- [13] Boashash, M. Mesbah, and P. Coldtitz, "Time-Frequency Detection of EEG Abnormalities," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, B. Boashash, Ed. Oxford, U.K.: Elsevier, 2003, pp. 663–670.
- [14] G. K. Wallace, "The JPEG still-picture compression standard," *Commun. ACM*, vol. 34, pp. 30–44, Apr. 1991.
- [15] Uwe Meyer-Baese, *Digital Signal Processing with Field Programmable Gate Arrays*, Springer-Verlag, 2001.
- [16] Robert D. Turney, Chris Dick, and Ali M. Reza, *Multirate Filters and Wavelets: From Theory to Implementation*, Xilinx Inc.
- [17] Spiliotopoulos, N.D. Zervas, C.E. Androurlidakis, G. Anagnostopoulos, S. Theoharis, Quantizing the 9/7 Daubechies Filter Coefficients for 2D DWT VLSI Implementations, 14th International Conference on Digital Signal Processing, pages 227 -231, vol.1, July 2002.
- [18] J.Ramirez, A. Garcia, U. Meyer-Baese, F. Taylor, P.G. Fernandez, A. Lloris, Design of RNS-Based Distributed Arithmetic DWT Filterbanks, *Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1193 -1196, vol.2, May 2001.
- [19] *Mathematical Methods and Algorithms for Signal Processing*, Todd K. Moon, Wynn C. Stirling, Prentice Hall, 2000
- [20] Bochner S., Chandrasekharan K. (1949), *Fourier Transforms*, Princeton University Press *Wavelet Transforms | A Quick Study*, Ivan W. Selesnick, Polytechnic University, Brooklyn, NY, September 27, 2007
- [21] Donoho, D.L.; I.M. Johnstone (1994), "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, Vol. 81, pp. 425–455.