INFORMATION TECHNOLOGY JOURNAL

Rate-distortion Estimation for Wavelet Video Coding Based on Multi-dimensional Gaussian Distribution

[1,2]Xuesong Jin, [1]Maoliu Lin, [2]Zhaojie Zhao and [2]Zhipeng Fan
[1]School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China
[2]School of Computer and Information Engineering, Harbin University of Commerce,
Harbin, China

Abstract: The rate allocation proposed a problem that exists in wavelet transformation based video coding. The goal of this paper is to design a solution for estimating the rate-distortion for wavelet subband faced by wavelet-based video coders. We proposed a new technique for establishing rate-distortion function using correlation between wavelet coefficients. In order to minimize the overall error of the reconstructed video, a rate allocation algorithm based on rate-distortion function was implemented. Experiment results show that the proposed rate-distortion model can predict the real rate-distortion performance of video encoder, can improve the reconstructed video quality significantly.

Key words: Wavelet video coding, optimal rate-allocation, rate-distortion function, multivariate normal distribution

INTRODUCTION

With the drastic developments in computer and network technologies, highly efficient compression and transmission for video information are always the primary targets. When develops a video Codec system, high coding efficiency, namely covering more information by using less bits, is not the only goal but also the adaptive ability to meet various of channel bandwidth, terminal capabilities and user demands. Thus in order to meet the new video coding challenges, Scalable Video Coding (SVC) was developed which provides the scalability of temporal, spatial and quality, etc. (Schwarz *et al.*, 2007).

The wavelet-based image compression algorithms, such as JPEG 2000 (Taubman and Marcellin, 2002), are known to outperform the discrete cosine transform (DCT)-based compression techniques in still image coding. It is natural to extend wavelet-based image processing techniques for video coding. The 3D wavelet transforms decompress video sequence into independent wavelet subbands. The embedded bit-stream is generated for each subband by performing the entropy coding. In order to produce the bit-stream with a given bit rate, the amount of bytes be allocated for each subband in the output embedded bit stream should be determined. This decision procedure raises a competition among all the subbands. Finally, the rate allocation is arbitrated by equal-slope criterion of rate-distortion (R-D) curve which plays a crucial role in rate allocation (Xu *et al.*, 2001).

According to the characteristic of the existing main R-D models, these R-D can be classified into three categories: analytic, empirical and semi-analytic (Xie *et al.*, 2009). Many R-D modeling methods in literatures have been proposed to estimate accurate R-D functions for different signals. Mallat and Falzon (1998) suggested a thorough analysis of R-D performance of wavelet transform coding in the low bit-rate region and presented an accurate model for still image coding. Their evaluation results revealed that the R-D function takes great different forms for low and high bit rate. Dai *et al.* (2004) combined Mallat's distortion calculation with the bit-rate estimation which is proposed by He and Mitra (2001), this approach achieve a close-form operational R-D model. Mingshi and Sachaar (2006) proposed a unified mathematical model to describe wavelet coefficient operational rate-distortion behavior based on their statistical investigation of a typical three-dimensional wavelet Codec. Their model can capture sequence characteristics accurately and reveal the relationship between wavelet decomposition levels and the overall R-D performance. Foo *et al.* (2008) analysed the encoding process and method of the encoder, thus established analytical models based on wavelet sub-band. This method is more complex. Meng-Ping and Nguyen (2010) constructed a linear distortion model, implemented the rate extractor based on optimize motion quality level. The PCRD-opt rate allocation method is main method in

Corresponding Author: Xuesong Jin, School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China

image coding which is also being used in wavelet-based video coding systems (Taubman and Marcellin, 2002).

Xu *et al.* (2001) obtained operational R-D point pairs, through utilizing the rate of the actual output of coder and the estimation distortion as quantization step decrease. They used the binary search algorithm to find the optimal bit-rate cut-off point. The proposed method estimated operational rate-distortion discrete points and rate allocation algorithm was implemented. This study presents a new method to establish the rate-distortion function of wavelet subband and then introduced a novel algorithm to optimize the rate allocation using the rate-distortion function.

THE ANALYSIS OF SUB-BAND COEFFICIENTS CHARACTERISTICS

There was several system schemes can be chosen in wavelet based scalable video coding system. This paper is based on T+2D structure. Motion Compensation Temporal Filtering (MCTF) was applied to remove interframe correlation along the time axis to implement temporal scalability. Then DWT is performed on the obtained temporal subband to finally generate T-S subband. We have found certain relationship between the variance and correlation of subband and the statistical property of original pixel. The relationship will be explained consequently.

The 3D wavelet video coding schemes is to perform MCTF (Ohm, 1994) on the input video sequence directly before spatial transform. The multi-level MCTF decomposes the video frames into several temporal subbands, such as temporal high-pass subbands and temporal low-pass subbands. After temporal decomposition, a spatial decomposition is applied to each temporal subband to further decompose the frames spatially. After temporal and spatial subband transforms, the coefficients will be coded with 3D Embedded Subband Coding with Optimal Truncation (3D-ESCOT) (Xu *et al.*, 2001). Basically, 3D-ESCOT follows the idea and coding procedure of EBCOT in JPEG 2000 (Taubman, 2000). The one-dimensional signal is adopted to illustrate because the convenience of analyzing subband coefficient characteristics. Supposed that signal is one-dimensional Gauss random process $\{x(n), n \in z\}$. Variance is σ^2, autocorrelation function of random process $R(k) = r^{|k|}\sigma^2$, $o \leq r \leq 1$, $k \in Z$. The signal decomposed with wavelet. The wavelet decomposition is equivalently two channel filter. The filter include of high-pass h_0 and low-pass h_1. As the result, subbands B_0 and B_1 are generated. As in Fig. 1,
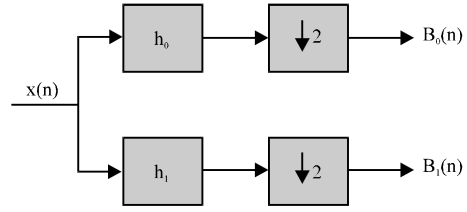

Fig. 1: One-dimensional signal wavelet decomposition

high-pass and low-pass subbands coefficients are obtained by following formula, respectively:

$$B_0(n) = \sum_i h_0(i)x_{2n+i} \tag{1}$$

$$B_1(n) = \sum_j h_1(i)x_{2n+j} \tag{2}$$

The variance of high-pass subband B_0 and low-pass subband $B_1(n)$ are depicted by $D(B_0)$ and $D(B_1)$, respectively. $D(B_0)$ and $D(B_1)$ are calculated from formula Eq. 3 and 4:

$$D(B_0) = D\left(\sum_i h_0(i)x_{2n+i}\right) = \sum_{each} h_0 h_0^T \otimes C \tag{3}$$

$$D(B_1) = D\left(\sum_j h_0(j)x_{2n+j}\right) = \sum_{each} h_1 h_1^T \otimes C \tag{4}$$

The notation $\sum$ pixels' summation of corresponding matrix and $\otimes$ means the element-multiply between corresponding matrix. Matrix C is as following:

$$C = E\left\{(X_m - E[X_m])(X_m - E[X_m])^T\right\} = \begin{pmatrix} 1 & r & \cdots & r^{m-1} \\ r & 1 & \cdots & r^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ r^{m-1} & r^{m-2} & \cdots & 1 \end{pmatrix} \sigma^2$$

$$E[B_0(j)B_0(k)] = E\left[\sum_l h_0(l)x_{2j+l} \sum_n h_0(n)x_{2k+n}\right] = \sum_l \sum_n h_0(l)h_0(n)r^{|2(k-j)+(n-l)|}\sigma^2 \tag{5}$$

$$E[B_1(j)B_1(k)] = E\left[\sum_l h_1(l)x_{2j+l} \sum_n h_1(n)x_{2k+n}\right] = \sum_l \sum_n h_1(l)h_1(n)r^{|2(k-j)+(n-l)|}\sigma^2 \tag{6}$$

It shows that the subband coefficients' variance is determined by filter coefficients and autocorrelation of original signal.

Although wavelet filter can remove the temporal and spatial correlation of video signal, there still are some relations between wavelet coefficients. Entropy coding

can utilize the relation to improve the coding efficiency. The autocorrelation of coefficients in sub-band B_0 and B_1 can be calculated from formula Eq. 5 and 6.

RATE-DISTORTION FUNCTION

The same frequency subband coefficients have similar statistical properties. Each subband, generated by temporal and spatial transforms, is divided into 3D coding blocks which will be coded independently, shown in Fig. 2. In the coding block, the correlation between the coefficients whose coordinate is (i_1, j_1, t_1) and (i_2, j_2, t_2) will be obtained by the following formula:

$$R\left((i_1,j_1,t_1),(i_2,j_2,t_2)\right)=\rho^{|i_1-i_2|+|j_1-j_2|+|t_1-t_2|} \tag{7}$$

In the coding block, the coefficients correlation can be approximated exponential function of the corresponding coefficients spatial location. The correlation of coefficient of coarser subbands is calculated by using Eq. 5 and 6, recursively.

For each coding block, bit-plane coding and context-based arithmetic coding are used. Coder encodes the current coefficients x exploiting the correlation between current coefficient and the adjacent coefficients in the position of the 3D coding block. Let $Nx = (x_u, x_d, x_l, x_r, x_f, x_b)$ denotes the context of current coefficients. The adjacent coefficients of current coefficient consist of up, down, left, right, before and behind coefficients, respectively, shown in Fig. 3. The coefficient x and the context Nx constitute a seven-dimensional random vector $X = (x, Nx)$ which obeys following Gaussian distribution:

$$f(X)=\frac{1}{(2\pi)^{7/2}|\Sigma|^{1/2}}e^{\left[-\frac{1}{2}X^T\Sigma^{-1}X\right]} \tag{8}$$

which:

$$\Sigma=\begin{pmatrix} 1 & \rho & \rho & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho^2 & \rho^2 & \rho^2 & \rho^2 & \rho^2 \\ \rho & \rho^2 & 1 & \rho^2 & \rho^2 & \rho^2 & \rho^2 \\ \rho & \rho^2 & \rho^2 & 1 & \rho^2 & \rho^2 & \rho^2 \\ \rho & \rho^2 & \rho^2 & \rho^2 & 1 & \rho^2 & \rho^2 \\ \rho & \rho^2 & \rho^2 & \rho^2 & \rho^2 & 1 & \rho^2 \\ \rho & \rho^2 & \rho^2 & \rho^2 & \rho^2 & \rho^2 & 1 \end{pmatrix}$$

In the covariance matrix Σ of different sub-bands, ρ can be obtained by Eq. 5 and 6.
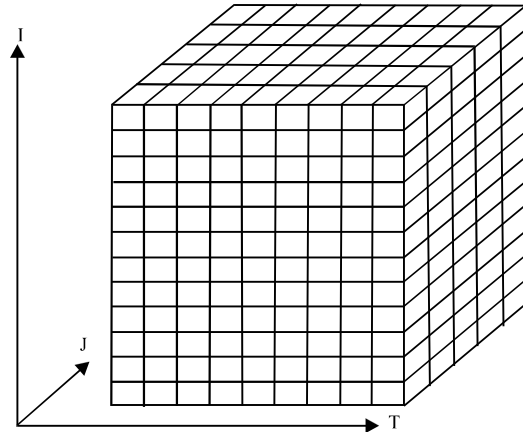

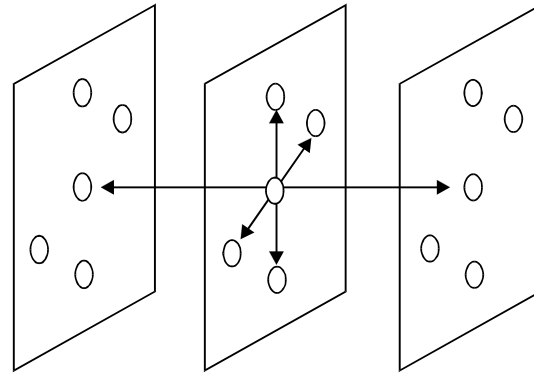Fig. 2: Three-dimensional wavelet coefficients code block

Fig. 3: The relationship between the current coding coefficients and the context correlation

Since the covariance matrix Σ is positive semi-definite, there is always a positive eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$. The corresponding eigenvectors $q_1, q_2, ..., q_n$ are easy to calculate. The eigenvectors constitute orthogonal matrix of Q. Covariance matrix Σ can be constituted by the following formula:

$$\Sigma=Q\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}Q^{-1}$$

and:

$$\Sigma^{-1}=Q\begin{pmatrix} \dfrac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \dfrac{1}{\lambda_{21}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \dfrac{1}{\lambda_n} \end{pmatrix}Q^{-1}$$

$$f(X) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}X^T Q \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_{21}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{pmatrix} Q^{-1}X\right] \quad (9)$$

Let $X' = Q^{-1}X$, the probability density function of vector X' becomes:

$$f(X') = \prod_{k=1}^{n}\frac{1}{\sqrt{2\pi\lambda_k}}\exp\left[-\frac{1}{2}\frac{(x_k')^2}{\lambda_k}\right] = f(x_1')f(x_2')\cdots f(x_n')$$

Each component of random vector X' is completely independent. Therefore, the entropy of random vector X' can be obtained by the following formula:

$$H(X) = H(X') = \sum_k H(x_k')$$

The same transformation is performed on Y, equally, $Y' = Q^{-1}Y$. The transformation will not affect average mutual information will not since the transform Q^{-1} is orthogonal transform and reversible. In addition, the Q is real matrix, so the above transformation is unitary transformations. Then, in the transformation, Euclidean distance remains the same length. The above transformation always uses the mean squared error to calculated distortion and ensures that the rate-distortion function of both sides is the equal.

In Eq. 9, each independent component x_k' of random vector X' obeys Gaussian distribution, whose variance is λ_k. The rate-distortion function of X' can be derived from the rate-distortion function of independent components. Let θ is the specified distortion parameters. The average distortion and bit rate of component can be calculated as follows:

$$D_\theta = \frac{1}{n}\sum_{k=1}^{n}\min(\theta, \lambda_k) \quad (10)$$

$$R_\theta = \frac{1}{n}\sum_{k=1}^{n}\max\left(0, \frac{1}{2}\log\frac{\lambda_k}{\theta}\right) \quad (11)$$

The values of rate and distortion will be calculated based Eq. 10 and 11 as the parameter θ varies (Berger, 1972). In this study, n is 7.

RATE-ALLOCATION ALGORITHM

The rate allocation problem for a video coder can be roughly stated as the determination of proper coding parameters so that the decoded video quality is optimized with respect to a certain fixed rate. After fully wavelet transformation, each sub-band is coded into bit-stream by entropy coder. The bit rate of each subband can be directly controlled to achieve the required distortion. For simplicity, rate-allocation is limited to a GOP. Let R_t be the rate assigned to the GOP. The rate allocation problem can be formulated as:

$$\min\{D_A\}, \text{subject to } R \le R_T \quad (12)$$

where, $D_A = \sum_i D_i$ and $R = R_\alpha + \sum_i R_i$, $R\alpha$ is the rate of coding motion information, such as macro-block partition mode, motion vectors, etc. Given the rate R_T for the GOP, we want to allocate the rate such that the overall distortion is minimized. There are two methods to solve the optimization problems of (12): Lagrangian optimization and dynamic programming. The Lagrange optimization method is widely used because of its simple and easy to implement. Restricted optimization problem used Lagrange multipliers can be converted into a non-limited form, cost function of construct as following:

$$J = D_A + \lambda \times (R - R_T) \quad (13)$$

where, λ is the Lagrange multiplier. Since D_A and R are monotone continuous curve of smooth, rate-distortion function is derivative everywhere. Solve the partial derivatives of cost function (13) and set partial derivatives to zero:

$$\frac{\partial J}{\partial R_i} = \frac{\partial D_i}{\partial R_i} + \lambda = 0 \quad (14)$$

For each sub-band, determined λ, you can obtain the all bit-rate R_i and distortion D_i. The coding parameter λ must be carefully selected so that condition $R \le R_T$ is guaranteed.

In fact, it is not easy to obtain the function expression of specific R-D function. We can get a number of R-D point pairs on the rate-distortion function. The bit allocation algorithm of using R-D point pairs can be described as follows:

- **Step 1:** Initialization: given a $\lambda_0 \le 0$, step size $d\lambda > 0$ and threshold $\varepsilon > 0$, make m = 0
- **Step 2:** For each temporal subband, calculate the first derivative of the rate-distortion function:

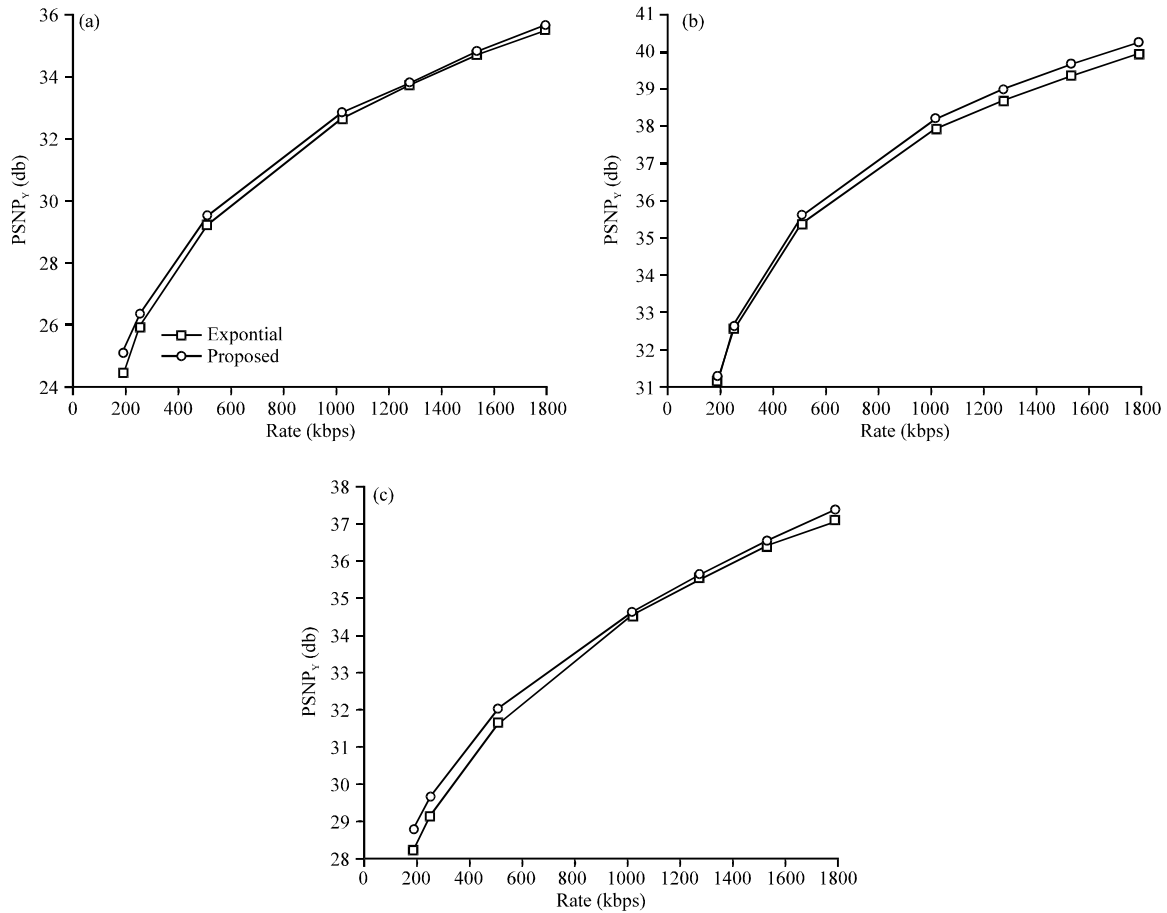$$\frac{\Delta D(R_i)}{\Delta R_i} = \lambda_m$$

Fig. 4: Comparing the test results, (a) Bus (b) Foreman and (c) Coastguard

- **Step 3:** Calculated and allocated the rate R_i of each temporal subband
- **Step 4:** Calculated $R = R_\alpha + \sum_i R_i$
- **Step 5:** If $(R_i - R)/R_T \leq \varepsilon$, then the algorithm will stop, otherwise, it will continue
- **Step 6:** Make $m = m+1$, $\lambda_m = \lambda_{m-1} + d\lambda$, transferred to the step 2)

EXPERIMENTAL RESULTS AND ANALYSIS

To verify the performance of the proposed rate-distortion function estimation algorithm, we choose the exponential rate-distortion model (Po-Yuen *et al.*, 1997; Taubman and Marcellin, 2002) for the temporal subband distortion. Then the temporal subband distortion is given by:

$$D_n = \sigma_n^2 2^{-\gamma_n R_n}$$

where, σ_n^2 is the source variance and γ_n is the coding efficiency parameter. For each temporal subband n, the coding efficiency parameter γ_n and the variance σ_n^2 have to be determined.

Video coding system platform based on discrete wavelet transform is built. The test sequences is CIF format with resolution size 352×288, frame rate is 30 fps. In the testing of the proposed method, the temporal decomposition level is set to four and using LG 5/3 filters. After decomposition, A GOP generates 16 sub-band in five types (one l^4 and h^4, two h^3, four h^2, eight h^1). Then, each temporal sub-band is decomposed using Daubechies 9/7 filters. The generated wavelet coefficients are coded into bit-stream by entropy Codec.

We compare the performance of proposed method to rate-distortion model of exponential function. The test results for the CIF sequences are plotted in Fig. 4. It can be seen from the Fig. 4, for the content rapid changes video sequence, such as bus, the result of the two algorithms is more consistent. For the moderate changes video sequence, such as foreman and coastguard, the PSNR of the decoded video has increased about 0.3 dB~0.5 dB. This means that using the proposed

rate-distortion model and rate allocation algorithm, the quality of the reconstruction image is improved compared to exponential model.

Owing to the calculation of correlation coefficient of each sub-band coefficients and diagonalization calculation of seven-dimensional covariance matrix, the computational complexity of proposed algorithm is higher.

CONCLUSION

A rate-distortion estimation algorithm based on multi-dimensional Gaussian distribution has been proposed for wavelet-based video coding system. Using the wavelet coefficient characteristic statistics, the encoder can derive the rate-distortion function. Experimental results show that this algorithm can be applied in wavelet video coding and achieve a better performance.

ACKNOWLEDGEMENTS

REFERENCES

Berger, T., 1972. Rate Distortion Theory: A Mathematical Basis for Data Compression. Prentice-Hall, Englewood Cliffs, USA., ISBN: 9780137531035.

Dai, M., D. Loguinov and H. Radha, 2004. Rate-distortion modeling of scalable video coders. IEEE Int. Conf. Image Process., 2: 1093-1096.

Foo, B., Y. Andreopoulos and M. van der Schaar, 2008. Analytical rate-distorition-complexity modeling of wavelet-based video coders. IEEE Trans. Signal Process., 56: 797-815.

He, Z. and S.K. Mitra, 2001. A unified rate-distortion analysis framework for transform coding. IEEE Trans Circuits Syst. Video Technol., 11: 1221-1236.

Mallat, S. and F. Falzon, 1998. Analysis of low bit rate image transform coding. IEEE Trans Signal Process., 46: 1027-1042.

Meng-Ping, K. and T.Q. Nguyen, 2010. Rate-Distortion optimized bitstream extractor for motion scalability in wavelet-based scalable video coding. IEEE Trans. Image Process., 19: 1214-1223.

Mingshi, W. and M. van der Schaar, 2006. Operational rate distortion modeling for wavelet video coders. IEEE Trans. Signal Process., 54: 3505-3517.

Ohm, J.R., 1994. Three-dimensional sub-band coding with motion compensation. IEEE Trans. Image Process., 3: 559-571.

Po-Yuen, C., J. Li, J. Kuo, 1997. Rate control for an embedded wavelet video coder. IEEE Trans. Circuits Syst. Video Technol., 7: 696-702.

Schwarz, H., D. Marpe and T. Wiegand, 2007. Overview of the scalable video coding extension of the H.264/AVC standard. IEEE Trans. Circuit Syst. Video Technol., 17: 1103-1120.

Taubman, D., 2000. High performance scalable image compression with EBCOT. IEEE Trans. Image Process., 9: 1158-1170.

Taubman, D.S. and M.W. Marcellin, 2002. JPEG2000: Image Compression Fundamentals, Standards and Practice. Vol. 1, Spronger, Boston, MA., USA., ISBN-13: 9780792375197, Pages: 773.

Xie, M., G. Wei and Y. Ling, 2009. Analysis and application on rate-distortion model oriented scalable video sequences. Inform. Technol. J., 8: 188-194.

Xu, J., Z. Xiong, S. Li and Y.Q. Zhang, 2001. Three-dimensional embedded subband coding with optimized truncation (3-D ESCOT). Applied Comput. Harmonic Anal., 10: 290-315.