A hybrid vector quantizer for enhanced image pyramid coding with application to volumetric image compression in confocal microscopy

Yegang Tao*[a], W. Paul Cockshott[a]
[a]Dept. of Computing Science, Univ. of Glasgow / University Ave., Glasgow, UK G12 8QQ

ABSTRACT

Three-dimensional image compression methods outperform their two-dimensional counterparts in the sense of higher rate-distortion performance for compressing volumetric image data. The state-of-the-art transform-based 3D compressors, such as 3D-SPIHT and 3D-DCT, are characterized for their rate control ability, where the qualities of the image, although are adjustable with respect to rates, are not explicitly controllable. A novel method, based on vector quantization in an enhanced image pyramid with error feedback, has been proposed, where the quality of the decompressed image only depends on the encoding of coefficients from the finest band and therefore a distortion-constraint transform coding is achieved. Compared to the previous image pyramid transform coders, its coding efficiency has been improved by using a cross-band classified vector quantizer (CBCVQ), where the encoding of current band will benefit from the encoding result from previous bands. Two explicit bit-allocation schemes, one is regarding the bit allocation across bands and the other is across the sub vector quantizers within each band, have been applied to minimize the total rate under the constraint of specified distortion. Evaluations have been performed on several data sets obtained by confocal laser scanning microscopy (CLSM) scans for vascular remodeling study. The results show that the proposed method has competitive compression performance for volumetric microscopic images, compared to other state-of-the-art methods. Moreover the distortion-constraint feature offers more flexible control than its rate-constraint counterpart in bio-medical image applications. Additionally, it effectively reduces the artefacts presented in other approaches at low bit rates and therefore achieved more subjective acceptance.

Keywords: Volumetric (3D) Image Compression, 3D Difference Pyramid, Vector Quantization, Bit Allocation, Confocal Laser Scanning Microscopy (CLSM)

1. INTRODUCTION

The development of digital imaging, especially the three-dimensional (3D) scanning in medical imaging, has increased image dimensions greatly, outstripping growth of storage media. Thus the gains to be had from compression are now more important than they were previously. Possible areas of application for compression of medical images have proliferated. They include wireless emergency medical services, battlefield and shipboard surgery and medicine, progressive browsing of databases, medical teaching archives and others. The distortion-rate performance of compression algorithms has improved enormously over the past decade [23], and a variety of volumetric image compressors have been proposed and investigated.

The first attempts to code a sequence of medical images were to hypothesize the collection of these images in time domain and methods similar to video coders, like MPEG, were used [12]. However it has been shown that that such algorithms are not adapted to the compression of 3D CLSM images because images are obtained as optical sections at different locations along the optical axis [19] and the motion compensation techniques cannot effectively capture the pixel-intensity correlations across consecutive frames [8]. Methods based on three-dimensional transforms, such as 3D-Discrete Cosine Transform (3D-DCT) [1, 17] and 3D-Discrete Wavelet Transform (3D-DWT) [16, 21], have received increased attentions. These real three-dimensional methods offer excellent coding efficiency by treating the whole sequence as a 3D volume data and capturing the inter-pixel correlations, no matter whether they are arranged in the same frame or across frames, within one procedure.

We noticed that in these state-of-the-arts 3D compressors, such as 3D-SPIHT [16] and 3D-DCT [21], although users can precisely control the bit rate or compression ratio, the image qualities are not explicitly controllable, where as in medical image applications the latter are more important than the former. In this case what an expected user concerns is the reconstructed image quality rather than the compressed image file size since the quality, describing how much

* tao@dcs.gla.ac,uk; tel 44 141 339-8855 ext. 0174; fax 44 141 330-3119; http://www.dcs.gla.ac.uk/~tao

Electronic Imaging and Multimedia Technology IV, edited by Chung-Sheng Li, Minerva M. Yeung, Proc. of SPIE Vol. 5637 (SPIE, Bellingham, WA, 2005)
0277-786X/05/$15 · doi: 10.1117/12.572972

information has lost, may affect their diagnostic results [3]. Consider a compression task like achieving target quality, it may be inefficient using the standard approaches. For instance specifying the encoder a target rate, which produces acceptable image quality on an image with large stationary distributions, may lead to unacceptable quality due to artefacts and loss of information on another image with lots of non-stationary distributions. Users have to manually balance the compression rate and image quality without explicit constraints from the encoder and therefore the compression must be performed multiple times to achieve target image qualities. Actually as it has been shown [24, 25] that the distortion measure, say mean-squared-error measures, between the original coefficients and their decoded versions in transform domain are not the same as that measured in the original image domain. Therefore these methods are inherently more like a rate-constraint system than a distortion-constraint system.

In this light, closed-loop image pyramids have been found attractive for their characteristics in strongly upper-bounding the reconstruction error due to the quantization error feedback scheme [2]. Compared to open-loop image pyramid, such as Burt's pyramid [4] and its variants [27, 14], they feed back the quantization error coming from previous pyramid level to the current level. The distortion of the decompressed image depends only on the quantization in the bottom level, and not on those in the other pyramid levels, and therefore the quality can be directly controlled by how well we encode the bottom level.

Since the image pyramid transform resembles subband decomposition, the coefficients in difference pyramids exhibit many statistical properties, among which the most significant characteristics are non-stationary intra-band distribution and similar distribution of inter-band coefficients along same orientation. A hybrid vector quantizer referred as CBCVQ was proposed and investigated. It can capture these dependencies by using a classification scheme for intra-band non-stationary distributions and a cross-band class decision rule for inter-band correlations.

In a closed-loop image pyramid the encoding across bands is dependent and the bit allocation problems should be investigated. In our method, it includes two aspects. One is the bit-allocation across pyramid bands or layers. In the sense of distortion-rate performance this can also be expressed as how we decide the quality profiles when encoding each pyramid level to minimize the total number of bits. The other aspect is how we allocate bits among sub quantizers to minimize the rate when encoding intra-band coefficients, under the constraint of distortion solved in the former problem.

This paper is organized as follows. In Section 2 an enhanced image pyramid transform is proposed where we improve the band decomposition performance by using an alternative to the conventional Burt's kernel; our three-dimensional image pyramid coder will then be introduced based on the enhanced kernel; the first bit-allocation problem will be discussed and the optimum solution to the quality profile for each pyramid level will be derived. In Section 3 a hybrid vector quantizer is introduced to exploit the dependencies among pyramid coefficients, the solution to the second bit-allocation problem will be derived. Finally, in section 4, we present coding results of coding of several CLSM image data sets, discuss and compare these with those of the state-of-the-art image volume coder 3D-SPIHT and 3D-DCT.

2. ENHANCED 3D IMAGE PYRAMID DECOMPOSITION

2.1. The Enhanced Image Pyramid Transform with Improved Band Filters

The motivation of the image pyramid transform is the decomposition of signals into several band components with minimal frequency overlap, so the original signal can be reconstructed from these components. It follows the basic theory in subband coding [29]. Consider an N-level pyramid decomposition, where subscript i denotes the pyramid level with $i=N-1$ for bottom level and $i=0$ for top level, the results after image pyramid transform are two pyramids. One is an image pyramid $\{IP_i\}$ consists a series of low-pass filtered half-resolution approximations, computed from next lower level IP_{i+1} using a low-pass half-band filter f_R and subsampling, of which the detailed features and the resolution is recursively reduced from the bottom level to the top. The other is a difference pyramid $\{DP_i\}$ consists of the coarsest approximation of the original IP_0 and a series of high-passed components corresponding to each layer in image pyramid, which are computed by taking the difference between the two consecutive layers, IP_i and IP_{i-1}, in image pyramid after upsampling IP_{i-1} to align the same number of coefficients as that in IP_i and removing the folding frequency due to upsampling using an interpolation filter f_E. This process can be formulated as,

$$\begin{cases} IP_i = \text{subsampling}(f_R(IP_{i+1})) \\ DP_i = IP_i - f_E(\text{upsampling}(IP_{i-1})) \end{cases} \quad (1).$$

An example of four-level (N=4) image pyramid decomposition is illustrated in Fig. 1. The efficiency of image pyramid transform is determined by the band-pass and band-stop performances of the band filter for reduction f_R and for

expansion f_E as discussed above. In ideal cases frequency responses of these two band filters would be rectangular shaped, where the decomposed components are orthogonal with each other without frequency overlapping. However, it is impossible to implement such ideal filter since it requires convolving with an infinite function – sinc(·). The 5-tap parametric Burt's kernel has been widely used in the literature [4, 14, 2] as an alternative to the ideal filter.

$$f_{Burt}^{(1)} = [0.25 - 0.5a, \quad 0.25, \quad a, \quad 0.25, \quad 0.25 - 0.5a] \tag{2}$$

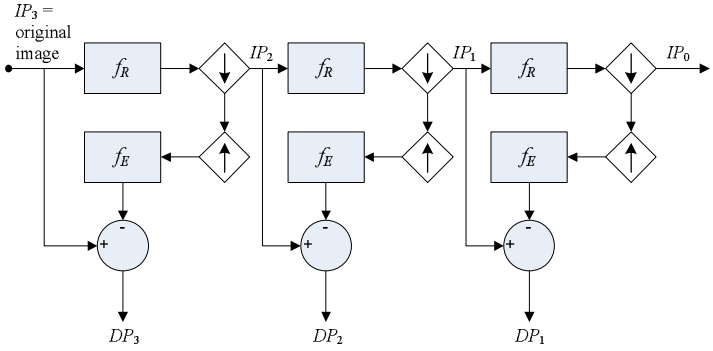

Figure 1: An example of four-level image pyramid decomposition.

It resembles a Gaussian-like kernel in both time domain and frequency domain and therefore the image pyramid generated using such kernel is often referred as Gaussian pyramid. Such Gaussian-like band-pass filters, although easy to implement and have very low computation complexity, are not well behaved in the sense of bandpass due to its shallow shape in the frequency domain. We therefore turned to derive the reduction filter f_R from an ideal sinc filter. Its general form $F_R^{(1)}(\cdot)$ is a sinc function $h(\cdot)$ multiplied by a same-size hamming function $\varphi(\cdot)$, as formulated in (2). An 11-tap halfband lowpass filter (Eq. 3) has been eventually exploited in our image pyramid transform. We made this decision based on the tradeoff between the performance and the computational complexity. Its impulse response and the spectrum is illustrated in Fig. 2. We also illustrates the Burt's kernel in the same figure for comparison purpose, where the generating parameter a=0.6 corresponding to the best performance in the sense of minimizing the equivalent entropy of the difference pyramid (refer to Eq. 11 in [2]).

$$F_R^{(1)}(n) = h^{(1)}(n) \cdot \varphi^{(1)}(n)$$
$$= \left\{ \frac{\sin(\frac{\pi}{2} i)}{\pi i} [0.54 - 0.46\cos(2\pi \frac{2i+n-1}{2n-2})], \quad i \in Z_{\left[-\frac{n-1}{2}, \frac{n-1}{2}\right]} \right\} \tag{3}$$

$$f_R^{(1)} = F_R^{(1)}(11)$$
$$= [0.0051, \ 0, \ -0.0422, \ 0, \ 0.2903, \ 0.5, \ 0.2903, \ 0, \ -0.0422, \ 0, \ 0.0051] \tag{4}$$

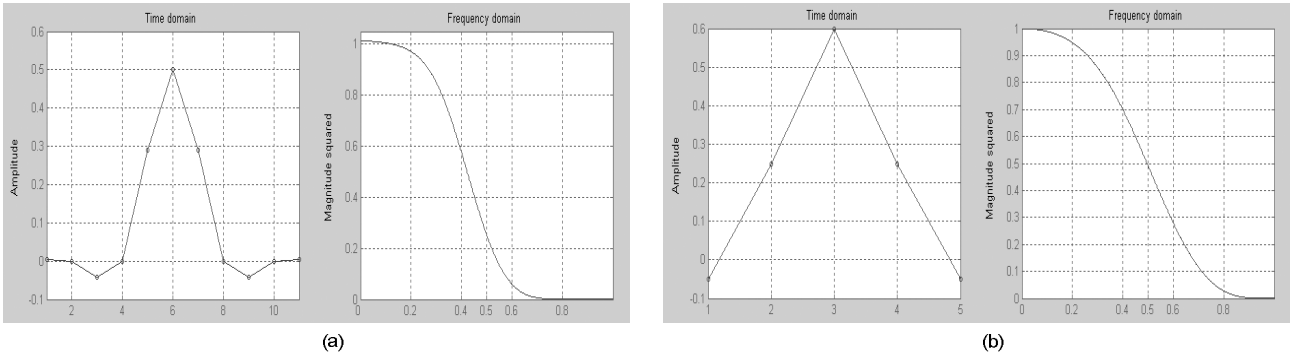

Figure 2: The impulse response and its spectrum of (a) the 11-tap filter $f_R^{(1)}$, compared to (b) the Burt's kernel $f_{Burt}^{(1)}$, with a=0.6.

Comparing with the Burt's kernel the 11-tap filter outperforms Burt's kernel since more energy will be stopped at cut-off frequency (0.5). People [28, 5] have shown that the failure to fully suppress the stopband, usually known as *frequency*

leakage, allows the offending frequencies to fold over into the passband range and tends to be more serious since they actually corrupt the components in the original signal and are visually perceived more readily. Therefore the folded high-frequency energy will eventually affect the difference pyramid and decrease the coding efficiency.

As to the interpolation filter f_E, the 7-tap halfband parametric kernel proposed in [2] has been considered. They demonstrated that this kernel has same computational cost as for 5-tap Burt's kernel, thanks to the null coefficients, but the performance is improved. We choose b=0.6 in our implementation, which has been verified in their works that this value would minimize the equivalent entropy of the difference pyramid.

$$f_E^{(1)} = f_{E,Aiazzi}^{(1)} = [0.5-b, \quad 0, \quad b, \quad 1, \quad b, \quad 0, \quad 0.5-b] \tag{5}$$

2.2. Three-dimensional Image Pyramid Coding with Quantization Noise Feedback

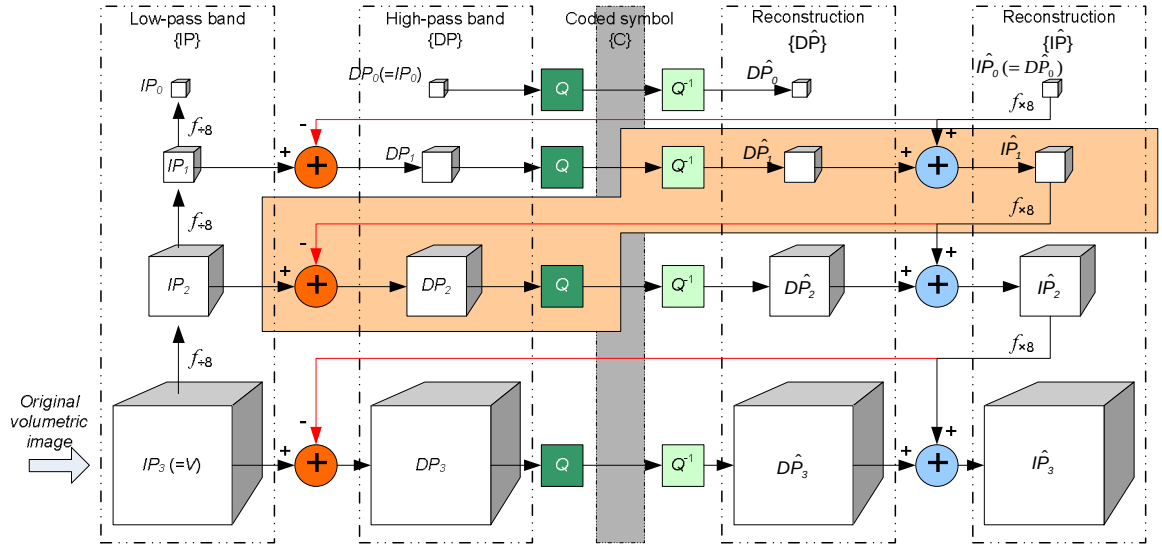

Figure 3: An example of encoding using four-level three-dimensional image pyramids with compression noise feedback.

An example of three-dimensional image pyramids coding for volumetric image data with compression noise feedback is illustrated in Fig. 3 and we generalized the coding process in Equation (5)-(8).

$$IP_i = \begin{cases} \text{Org.Img.} & i = N-1 \\ f_{\downarrow 8}(IP_{i+1}), & 0 \le i < N-1 \end{cases} \tag{5}$$

$$DP_i = \begin{cases} IP_i, & i = 0 \\ IP_i - f_{\uparrow 8}(\hat{IP}_{i-1}), & 0 < i \le N-1 \end{cases} \tag{6}$$

$$\hat{DP}_i = Q^{-1}(Q(DP_i)) \tag{7}$$

$$\hat{IP}_i = \begin{cases} \hat{DP}_i, & i = 0 \\ f_{\uparrow 8}(\hat{IP}_{i-1}) + \hat{DP}_i, & 0 < i \le N-1 \end{cases} \tag{8}.$$

We extended the one-dimensional reduction kernel $f_R^{(1)}$ and the expansion kernel $f_E^{(1)}$ to three dimensions by computing the cross-product of three identical one-dimensional band filters,

$$f_R = f_R^{(1)} \otimes f_R^{(1)} \otimes f_R^{(1)} \tag{9}$$

and

$$f_E = f_E^{(1)} \otimes f_E^{(1)} \otimes f_E^{(1)} \tag{10}.$$

Due to the separability of f_R and f_E, the convolution with the three-dimensional filter can be computed as three individual convolution with one-dimensional filter, and thus reducing the computation cost greatly. The reduction operator $f_{\downarrow 8}$ consists of a lowpass filtering using f_R followed by subsampling in each dimension; while the expansion operator consists of an up-sampling operation followed by interpolation filter f_E.

As we can see in Fig. 3 the quantization error from next upper level is presented in the reconstructed difference pyramid and eventually in the reconstructed image pyramid (the light orange part). Therefore using the reconstructed image pyramid to compute the difference will feed the error from previous encoding state back to current encoding state. Such error-feedback scheme can guarantee a flexible control of decompressed image quality. Consider encoding a four-level image pyramid as illustrated in Fig. 3 $I\hat{P}_3$ is the reconstruction version of IP_3, which is the original image. The difference between them is a kind of quality measure decompressed images. By rewriting (6), we will have:

$$\begin{aligned} & D(IP_3, I\hat{P}_3) \\ & = D([DP_3 + f_{\times 8}(I\hat{P}_2)], I\hat{P}_3) = D([DP_3 + f_{\times 8}(I\hat{P}_2)],[D\hat{P}_3 + f_{\times 8}(I\hat{P}_2)]) \\ & = D(DP_3, D\hat{P}_3) \\ & = D(DP_3, Q^{-1}(Q(DP_3))) \end{aligned} \tag{11}.$$

Therefore the decompressed image quality is only determined by how well we encode DP_3 and the quality can be controlled if we have designed a coder Q with distortion constraint.

2.3. Rate-minimizing Cross-band Distortion Control

The problem regarding the bit allocation among layers in a closed-loop image pyramid has been investigated [13, 2, 20]. Among these works, they describe this problem in the context of pyramid coding as determining the average number-of-bit per sample (or rate) r_i for each level which would minimize the distortion D between the reconstructed image and the original while under the constraint that the overall rate R is no more than a predefined value R_0,

$$\text{optimum rate scheme} : \operatorname{argmin}\left[\sum_{i=0}^{N-1} \delta_I^{(d)}(r_i)\right]\Big|_{\sum r_i < R_0} \tag{12},$$

where $\delta_i^{(d)}(\cdot)$ is the rate-distortion function corresponding to the encoding of coefficients in i^{th} level. Horn [13] found the solution a closed-loop pyramid, and they described it in the following form,

$$r_i = \begin{cases} \dfrac{1}{2}\log 2 \dfrac{\alpha\sigma_i^2 (M_{i-1} - M_i)}{\sigma_{i-1}^2 (M_i - M_{i+1})} & i > 0 \\ R_0 - \dfrac{\sum_{i=1}^{N-1} M_i r_i}{\sum_{i=0}^{N-1} M_i} & i = 0 \end{cases} \tag{13},$$

where M_i and σ_i^2 are the number of samples and their variance, in the i^{th} level of difference pyramid in the opened-loop form as illustrated in Fig. 1, respectively; α is referred as the power transfer factor indicating how much energy from previous level has been propagated to the current level, which is only determined by the interpolation filter f_E.

Their works cannot be directly applied in our case since the vector quantizers we designed were not explicitly determined by the rates allocated to each pyramid level. Alternatively we solved the bit-allocation problem with respect to distortion control scheme rather than conventional rate control schemes. Therefore the problem in our case is to determine the distortion profile d_i for encoding each level of pyramid which would minimize the overall rate R while under the constraint that the final reconstructed distortion D is no more than a predefined value D_0. Compared to the optimum rate scheme, we can describe the problem of optimum distortion scheme as,

$$\text{optimum distortion scheme} : \operatorname{argmin}\left[\sum_{i=0}^{N-1} \delta_i^{(r)}(d_i)\right]\Big|_{d_{N-1} < D_0} \tag{14},$$

where $\delta_i^{(r)}(\cdot)$ is the distortion-rate function for coefficients in level i. Thanks to the inter-changeability between $\delta_i^{(r)}(\cdot)$ and $\delta_i^{(d)}(\cdot)$, and follow the same assumption in [13, 2] that the samples in difference pyramid can be modeled as a zero-mean Laplacian distribution such that,

$$d_i = \delta_i^{(d)}(r_i) = \sigma_i^2 2^{-2r_i} \tag{15}.$$

Considering the power gain interpolation filter (Eq. 4), which can be computed according to [2] as

$$\alpha = (2b^2 - b + 3/4)^3 \tag{16},$$

we substituted (15) and (16) into (13) and then derived the solution for (14) as,

$$d_i = (\sigma_i^2 + d_{i+1})\frac{\sigma_{i-1}^2(M_i - M_{i+1})}{\alpha\sigma_i^2(M_{i-1} - M_i)} = \frac{(\sigma_i^2 + d_{i+1})\sigma_{i-1}^2}{8\alpha\sigma_i^2} \cong \frac{(\sigma_i^2 + d_{i+1})\sigma_{i-1}^2}{5.27\sigma_i^2}, \quad i > 0 \qquad (17).$$

Now we can find out the optimum distortion scheme for each layer with four steps: (1) building the opened-loop difference pyramid according to Fig. 1; (2) specifying the distortion d_0 for level 0, which would determine the final reconstructed quality; (3) specifying the distortion d_{N-1} for top level (in our implementations, we always set d_{N-1}=0); and finally (4) computing the distortion profile for other layers, from N-2 to 1, using Eq. 17.

3. A CROSS-BAND CLASSIFIED QUANTIZER

Because of the subband decomposition properties of image pyramid transform and the non-stationary distribution of image pixels, coefficients in the difference pyramid are correlated. There are two types of dependencies which could be considered and exploited by the encoder to improve the coding efficiency. One is the non-stationary distribution of intra-band coefficients; and the other is the similar distribution of inter-band coefficients crossing bands along the same directions. A hybrid vector quantizer has been proposed. The intra-band coding is based on classification according to the statistical properties of samples in each level, where an intra-band vector is firstly associated with a sub-VQ by the corresponding classification map, and then this sub-VQ is used to encode the vector based on a nearest-neighbor mapping rule. The inter-band coding is achieved by building the current-level classification map based on the encoding information from previous levels, and therefore the side information for classification map don't need to be explicitly transmitted to the decode side, but can be derived when the decoding of the previous levels is finished.

3.1. Quantizer Design

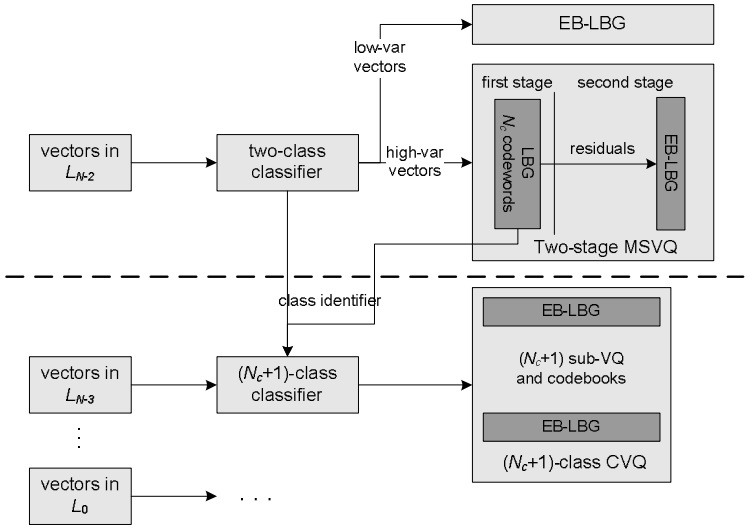

Figure 4: The block diagram of the cross-band classified VQ encoding.

From Fig. 4 we can see that the cross-band encoding can be basically divided into two parts, as indicated with the dashed line. The first part (above the dashed line) includes the encoding of vectors in level N-2, the first high-pass band in the difference pyramid. Moreover its intermediate outputs will be used to derive the classification maps for encoding its lower levels. The second part (below the dashed line) consists of several classified-VQ encoders with each for every level below N-2. The overhead bits of classification information of vectors from these levels are not explicitly transmitted to the decoder, but implicitly obtained from the same intermediate outputs in part one.

Such a cross-band encoding scheme has been investigated in [6], where they encoded L_{N-2} using a LBG-VQ [18] with codebook size of N_C to generate the classification map thanks to its clustering property, so the vectors from L_{N-2} will be grouped into N_C classes. However we found that the choice of N_C has a significant impact on the coding efficiency. A large value of N_C, like 256 in [6], will improve the reconstructed quality, but the classification efficiency will decrease since sometimes the vectors would be 'over-classified' which lead to too few vectors in a class. Therefore we introduced a two-stage multi-stage vector quantizer (MSVQ) [15], in which a LBG-VQ with small codebook size, e.g. 8 or 16, was used in the first stage. It will on one hand produce the residual signals for the second-stage encoding; and on the other

hand generate indices associated with each vector, which can be used to derive classification map of the following levels. Moreover, we have observed that the LBG-VQ in the first stage in MSVQ doesn't achieve good classification gain for low-variation vectors, which would expect to contain a major part of samples than the high-variation vectors in the difference pyramid. Therefore a two-class classifier was preceded before MSVQ and thus a vector is firstly characterized as a low-variation vector or a high-variation vector by comparing its variance with a preset threshold T_v, then MSVQ is used for high-variation vectors only and a LBG-VQ for low-variation vectors. The outputs from the two-class classifier together with the indices from the LBG-VQ in the first stage in MSVQ would be used as the class identifiers for the following levels.

After the encoding on level N-2, each vector in this level will be associated with a class identifier, either from the two-class classifier or the indices from the LBG-VQ in the first stage in MSVQ, which would be used for the following levels. Consider a vector with size $v_x \times v_y \times v_z$ at the location (x, y, z) in level N-2 is associated with class c_i, we apply the cross-band classification rule on level i ($i \geq N$-3) such that samples, within block with size $\tau v_x \times \tau v_y \times \tau v_z$ ($\tau = 2^{N-2-i}$, the scaling factor between level i and N-2) at the location $(\tau x, \tau y, \tau z)$corresponding to the same orientation as that in level N-2, will be associated with class c_i. An example of this classification rule on level N-3 is illustrated in Fig. 5.

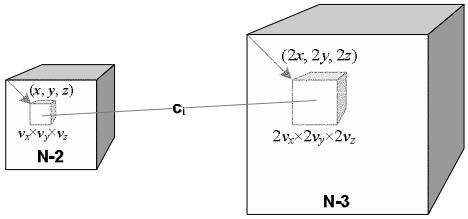

Figure 5: The cross-band classification rule applied on level N-3, the next lower level to N-2.

Once we obtained the classification map for every level below N-2, the encoding on these levels would be similar as a ordinary classified-VQ. One should notice that the vector size on these levels doesn't need to be the same as the block size for classification, but the block size has to be in multiple with vector size to make sure that samples within a vector are all belong to a unique class. Actually a similar vector forming technique as proposed in [22] was exploited, where we modified the decision rule to account for the classification such that the vector shape is determined by the variance of samples belonging to class c_i along each direction, not by that of the entire intra-band samples in previous work.

3.2. Distortion-constraint Intra-band Bit Allocation

It has been recognized that the distortion profile for each level is pre-determined before intra-band encoding, and thus the quantization on each level is constrained by the corresponding distortion. The sub-VQs associated with each class are basically LBG quantizers but under the distortion constraint and will be referred as distortion-constraint LBG (DC-LBG). Consider there are C sub-VQs, for example two for level N-2 and N_c+1 for other levels, designed as DC-LBG, the distortion profile for the i^{th} level, d_i, doesn't explicitly specify the maximum distortion $d_{i,j}$ allowed for each sub-VQ, but they should satisfy that the average distortion is no greater than the distortion profile:

$$\frac{\sum_{j=1}^{C} M_{i,j} d_{i,j}}{\sum_{j=1}^{C} M_{i,j}} \leq d_i \qquad (18),$$

where $M_{i,j}$ denotes the number of sample in level i belonging to class j. Therefore the bit allocation problem in this case is to minimize the average rate for encoding level i while under the constraint as Eq. 18. Since the encoding of each sub-VQ is independent we can find the solution based on the 'equal-slope' method, which is well-known in economics as 'Pareto Optimality' [26] and widely used in compression applications [7]. Its intuitive explanation can be briefly described as follows: Suppose you are operating on a variety of quantizers with each individually encoding a part of a source without overlapping. Each quantizer, which operates at different rates, yields a spread of possible distortions and rates for its corresponding sub-source. If the channel allows you to increase your overall rate slightly, which quantizer should you choose to increase? The one that produce maximum normalized distortion reduction among these quantizers would be chosen, which means its normalized rate-distortion curve at current rate profiles has the maximum magnitude of slope among others'. Consider a $k_{i,j}$-dimensional sub-LBGVQ $q_{i,j}$ is designed at rate λ, its normalized magnitude of slope $\varepsilon_{i,j}$ at rate λ can be computed by taking the difference of distortion of $q_{i,j}$ at rate λ and λ+1, corresponding to the codebook size 2^{λ} and $2^{\lambda+1}$, respectively; and then normalized by its percentage in the entire samples in this level:

$$\varepsilon_{i,j}(\lambda) = \frac{M_{i,j} \times k_{i,j} \times [\delta_{i,j}^{(d)}(\lambda) - \delta_{i,j}^{(d)}(\lambda+1)]}{\sum_{j=1}^{C} M_{i,j}} \quad (19).$$

Algorithm 1: Optimum intra-band bit allocation for level i under overall distortion constraint

- Step-1 (Initialization) Initialize each sub-VQ at rate $\lambda_{i,j}$=0 with codebook size = 1. Compute $\varepsilon_{i,j}(\lambda_{i,j})$ for each sub-VQ according to Eq. 19. Store $\varepsilon_{i,j}(\lambda_{i,j})$ in an one-dimensional array E with E[j]=$\varepsilon_{i,j}(\lambda_{i,j})$.
- Step-2 (Stopping rule) Compute the overall average distortion d_i' as

$$d_i^{'} = \frac{\sum_{j=1}^{C} M_{i,j} \times k_{i,j} \times \delta_{i,j}^{(d)}(\lambda_{i,j})}{\sum_{j=1}^{C} M_{i,j}} \quad (20).$$

  If d_i'<d_i we continue to Step-3; otherwise stop, with each sub-LBGVQ $q_{i,j}$ designed at rate $\lambda_{i,j}$.
- Step-3 (Find the sub-VQ with the maximum magnitude of slope) Search E to obtain the index $j_{\max}$ corresponding to the maximum magnitude of slope produced by $q_{i,j}$ at rate $\lambda_{i,j}$.
- Step-4 (Allocate bit to the sub-VQ $q_{i,j_{\max}}$) Re-build the sub-VQ $q_{i,j_{\max}}$ by doubling its codebook size from $2^{\lambda_{i,j_{\max}}}$ to $2^{\lambda_{i,j_{\max}}+1}$. Update its rate as $\lambda_{i,j_{\max}} = \lambda_{i,j_{\max}} + 1$.
- Step-5 (Update E and iteration) Replace E[$j_{\max}$] with $\varepsilon_{i,j_{\max}}(\lambda_{i,j_{\max}})$. Go back to Step 2.

4. RESULTS AND DISCUSSION

4.1. Image Acquisition

The coding performance of the proposed coder is evaluated on three data sets obtained from CLSM device. They are captured by the EC FP5 partnership (VASCAN2000) for vascular imaging and vascular remodeling studies [3, 10]. Each data set has typical statistical characteristics that are different from others, and therefore they were chosen as prototypes of CLSM images for the purpose of evaluating compression performance. For instance B3CGP and C5 data sets are characterized as high-variation distribution and detailed features, while Hum_Glt_C data set has many smoothed and low-variation regions. All the three groups of data sets are captured with Fluorescent-CLSM, either optical laser-scanning confocal microscopy or multi-photon laser-scanning microscopy. Specimens are tissues from human gluteal arteries and rat mesenteric arteries. They are fixed on the stage such that the imaging plane is parallel to the direction of blood flow. Specimens are stained with single type of fluorophore (single-labeled) and therefore images have one color channel and generated as 256 gray levels for each pixel, or 8 bits per-pixel. The imaging properties for each data set are listed in Table 1.

Table 1: Imaging properties for each data set.

Data set labels[1]	Description	Image acquisition specifications	Image dimensions (pixel)	Pixel dimensions (μm/pixel)
B3CGP	rat mesenteric artery stained with fluorescent BODIPY-CGP12177 to illustrate the distribution of cells and nerves	ImT[1]: Confocal Obj[2]: ×40	X=Y= 256 Z(frames) = 178	X=Y= 0.98 Z= 0.35
C5	rat mesenteric artery stained with propidium iodide dyes (nuclear dyes) to illustrate the distribution of nuclei of adventitial cells, smooth muscle cells and endothelial cells	ImT: Two-photon Obj: ×20	X=Y= 256 Z= 147	X=Y= 1.0 Z= 1.0
Hum_Glt_C	human gluteal artery with propidium iodide dyes (nuclear dyes) to show the shape of nuclei of the same three types of cells	ImT: Two-photon Obj: ×20	X=Y= 512 Z= 82	X=Y= 0.6 Z= 1.0

[1]: ImT: imaging technique used for capturing - optical laser-scanning confocal microscopy or multi-photon laser-scanning microscopy.
[2]: Obj: objective magnification factor.

4.2. Objective Evaluations

The most important feature of the proposed method is the precise control of decompressed image quality. Therefore in our method what a user specified is the expected quality or distortion rather than the expected rate or compression ratio

in those rate-parameterized compressors. Moreover high distortion-rate performances are achieved thanks to the image pyramid transform, inter-/intra-band quantization and optimum bit allocation schemes. For instance consider a user-specified profile as MSE=40, the encoding process can be completed in the following steps: (1) building an *N*-level image pyramid and the difference pyramid according to Fig. 1 with filters specified by Eq. 4 and 5; (2) specify the distortion for bottom level as d_0=40; (3) find out the distortion profiles for other layers according to Eq. 17; (4) Recursively building and encoding the difference pyramid from top to bottom according to Fig. 3 and Eq. 5~8; (5) individually encoding intra-band coefficients from level N-2 to level 0 using the cross-band classified vector quantizer and intra-band bit-allocation scheme (Algorithm 1); (6) and finally gradually allocating bits (Algorithm 1) to level 0 to decrease the distortion until its MSE measure becomes less than 40.

With the distortion-parameterized algorithms in our method the distortion-rate performance, is computed by measuring at four different quality profiles: PSNR=30dB, 33dB, 35dB and 40dB, which corresponding to MSE=65.54, 32.85, 20.72 and 6.55 respectively. The level of image pyramid has been considered such that the image size on top level is 32×32 in x-y plane. So we built a 4-level pyramid for B3CGP and C5 data sets; while a 5-level pyramid for Hum_Glt_C data set. The codebook size for the first stage MSVQ N_c is 16 and the high-variation/low-variation threshold T_v is 30. As we discussed above a small value can not well capture the non-stationary property of intra-band signal while a large value may lead to an over-classified effect, both will decrease the coding efficiency. Actually these two parameters are signal dependent, here, however, we didn't consider their adaptabilities to different data sets and we empirically choose N_c=16 and T_v=30 for all the data sets.

A comparison study has been performed by comparing our distortion-constraint method with the state-of-the-art 3D image coder 3D-SPIHT [16] and 3D-DCT [21]. We noticed that these methods are basically rate-parameterized and don't have distortion-control functionality. The performance at above PSNRs were obtained by first computing their rate-distortion curves with respect to six different rates, then find the rates on these curves which are corresponding to the desired PSNRs and finally delicate adjustments around these rates to make the actually PSNRs approach to the desired values as much as possible, for instance falling into the range of $\pm 2\%$ of the expected value. Table 2 lists the actual PSNRs and corresponding rates for three methods measured at four desired PSNRs for each data set. We also indicated the number of encoding attempts performed before the compressor can output the desired distortion, where for rate-parameterized methods it consisted of the average number of attempts to compute the rate-distortion curve and the number of attempts to refine the output.

Table 2: Distortion-rate measures of three methods for each data set.

Data set	Expected PSNR (dB)	Distortion-constraint Pyramid Coder		3D-SPIHT		3D-DCT	
		Actuall PSNR (No. of Attempts)	Rate (bpp)	Actuall PSNR	Rate	Actuall PSNR	Rate
B3CGP	30	30.00 (1)	0.12	29.96 (3.5)	0.13	30.07 (4.5)	0.21
	33	33.00 (1)	0.25	33.09 (2.5)	0.27	33.02 (4.5)	0.39
	35	34.87 (1)	0.39	35.03 (2.5)	0.42	34.97 (3.5)	0.58
	40	40.10 (1)	1.04	40.02 (3.5)	1.08	39.96 (3.5)	1.33
C5	30	29.97 (1)	0.08	29.97 (3.5)	0.09	29.99 (3.5)	0.11
	33	33.00 (1)	0.16	33.01 (2.5)	0.18	33.02 (3.5)	0.23
	35	35.02 (1)	0.28	35.08 (2.5)	0.30	34.93 (3.5)	0.34
	40	39.98 (1)	0.90	40.01 (3.5)	0.95	40.18 (3.5)	1.01
HUM_GLT_C	30	30.30 (1)	0.026	30.31 (4.5)	0.03	30.09 (4.5)	0.06
	33	33.16 (1)	0.05	33.18 (4.5)	0.06	32.95 (3.5)	0.10
	35	34.91 (1)	0.083	34.88 (3.5)	0.09	35.56 (4.5)	0.16
	40	39.80 (1)	0.263	40.04 (4.5)	0.27	40.37 (4.5)	0.38

From Table 2 we found that in the distortion-constraint pyramid coder the overall distortion can be effectively controlled by the closed-loop pyramid structures and the distortion-constraint quantizer and bit-allocation algorithm, where the PSNRs of the actual outputs are restricted within the range of $\pm 1\%$ of the expected values. So the decompressed image quality can be directly controlled in our method. However it is a different case in 3D-SPIHT and 3D-DCT, where the control of distortion is implicit and actually it is achieved by pre-computing the rate-distortion curve and delicately controlling the bit rates. For instance 3D-SPIHT averagely needs 3.5 attempts of pre-computation before it reaches the target distortion, and the attempts are even more in 3D-DCT since its bit-allocation scheme is not as accurate as that in

3D-SPIHT. Therefore to meet the requirement like compressing at the desired image quality, our distortion-parameterized method significantly outperforms the rate-parameterized methods in the sense of flexibility and simplicity.

Figure 6 illustrates the distortion-rate curves of three methods for each data set. We can see that our pyramid coder has comparable performances to the state-of-the-arts methods for volumetric images having different statistical characteristics. Moreover the advantages of our method become more significant at moderate qualities, like PSNR=33~35. This is consistent with the reports from previous works [22] as thanks to the error feedback scheme in image pyramid structures and the high rate-distortion performance of vector quantization at low rates [11].

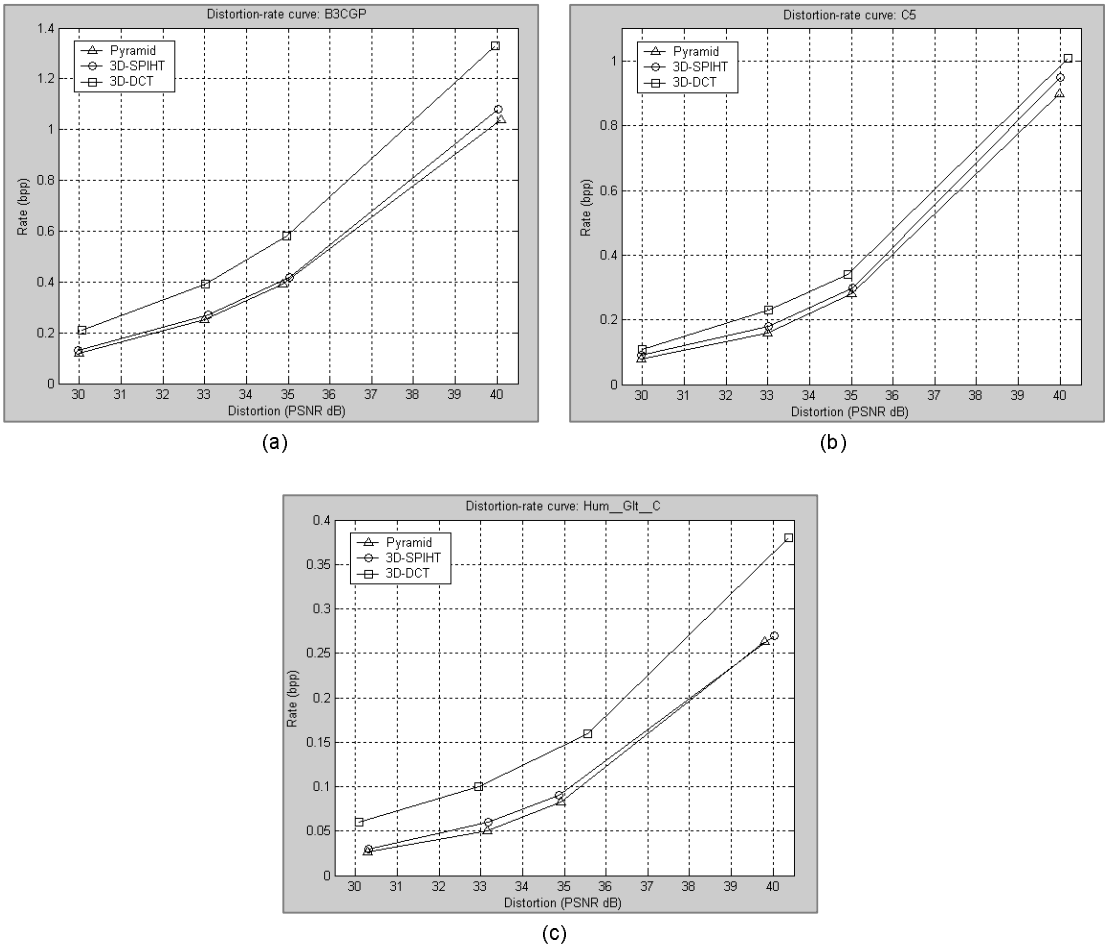

Figure 6: Distortion-rate performances of three methods, Distortion-constraint pyramid, 3D-SPIHT and 3D-DCT, evaluated on (a) B3CGP, (b) C5 and (c) Hum_Glt_C data sets, respectively.

Additionally from the decompressed images we noticed that these methods suffered from different types of artifacts and presented different visual quality at same objective measurement, say MSE, especially at low bit rates. Fig. 7 illustrates the 31st frame from Hum_Glt_C data set compressed at quality PSNR=30 by three methods respectively. Both the pyramid method and the 3D-SPIHT present blurring artifact, moreover unexpected high frequency components are apparent in 3D-SPIHT within the smooth regions around the nuclei. As to 3D-DCT the shapes of the interested regions, such as nuclei, are distorted by the blocking artifacts, which as discussed in [8] may worse to be acceptable in vascular remodeling studies. Therefore as stated by [9] objective measurement is not enough to qualify the decompressed image quality in medical applications. More issues regarding the subjective measurements and consistency tests, whose purposes are investigating the effect of compression approaches on results of the diagnostic tasks, including measurement of structures, detection of lesions and interpretation of texture, should be considered. They will be investigated in our future work.

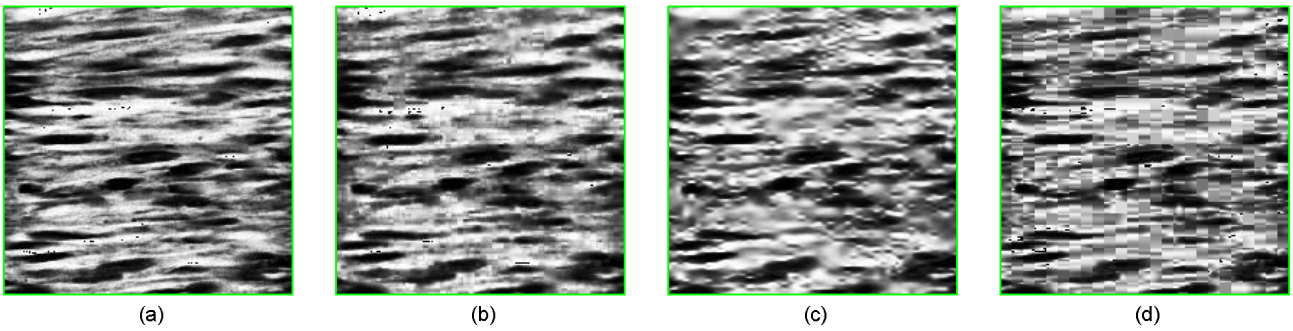

Figure 7: Coding results of Hum_Glt_C data set. A region of 200×200 from the 31st frame is illustrated. (a) The original image; (b) Compressed by Distortion-constraint pyramid; (c) 3D-SPIHT and (d) 3D-DCT. The pixel intensities are inversed for better visualization.

5. CONCLUSION

It has been indicated that the control of image quality plays a more important role in medical image application than that of compression ratio, where most of the state-of-the-art methods are rate-parameterized and cannot achieve target quality directly. We investigated this problem and proposed a novel method based on closed-loop image pyramid and vector quantization. The solution is simple and explicit thanks to the error feedback scheme in image pyramid structures and the distortion-constraint vector quantization on each pyramid layer. The evaluation results on several microscopic volumetric data sets show that our method is effective in flexible control of decompressed image quality. The distortion-rate performances have been improved by the using of enhanced image pyramid transform, cross-band classified vector quantization and optimum inter-band and intra-band bit allocation schemes. They are comparable to 3D-SPIHT and 3D-DCT. Moreover we noticed that the visual qualities of the decompressed images may be different and depends on the signal and the compression algorithm given the same objective criterion; and our method present more acceptable visual quality due to its effective reduction of artefacts.

REFERENCES

1. G. P. Abousleman, M. W. Marcellin and B. R. Hunt, *Compression of hyperspectral imagery using the 3d dct and hybrid dpcm/dct*, ieee tRans. on Geosci. Remote Sensing **33** (1995), 26-34.
2. B. Aiazzi, L. Alparone, S. Baronti and F. Lotti, *Lossless image compression by quantization feedback in a content-driven enhanced laplacian pyramid*, IEEE Trans. on Image Processing **6** (1997), 831-843.
3. S. M. Arribas, C. Daly and I. C. McGrath, "Measurements of vascular remodeling by confocal microscopy," *Confocal microscopy (methods in enzymology, volume 307)*, P. M. Conn (Editor), Academic Press, 1999, pp. 246-273.
4. P. J. Burt and E. H. Adelson, *The laplacian pyramid as a compact image coder*, IEEE Trans. on Commun. **31** (1983), 532-540.
5. K. R. Castleman, *Digital image processing*, Prentice Hall, 1995.
6. R.-F. Chang and Y.-L. Huang, *Finite-state vector quantization by exploiting interband and intraband correlations for subband image coding*, IEEE Trans. on Image Processing **5** (1996), no. 2, 374-378.
7. W. P. Cockshott, A. Cottrell and D. Breslin, *Optimal allocation of bits to frames in compressed video*, SCI 2000, Academic Press, 2000, pp. 249-255.
8. W. P. Cockshott, Y. Tao, G. Gao, P. Balch, A. M. Briones and C. Daly, *Confocal microscopic image sequence compression using vector quantization and 3d pyramids*, SCANNING-The Journal of Scanning Microscopies **25** (2003), no. 5, 247-256.

9. P. C. Cosman, R. M. Gray and R. A. Olshen, *Evaluating quality of compressed medical images: Snr, subjective rating, and diagnostic accuracy*, Proc. of IEEE **82** (1994), no. 6, 919-932.

10. C. Daly, "Receptor imaging," *VASCAN 2000 Tech. Rep. 2*, 2002.

11. R. M. Gray and D. L. Neuhoff, *Quantization*, IEEE Trans. on Inform. Theory **44** (1998), 2325-2383.

12. B. K. T. Ho, P. Saipetch, J. Wei, M. Ma, J. Villanseno and M.-J. Tsai, *Video compression algorithm for dynamic angiographic images*, Proc. SPIE Medical Imaging, 1994, p.^pp. 302-309.

13. U. Horn, T. Wiegand and B. Girod, *Bit allocation methods for closed-loop coding of oversampled pyramid decompositions*, Proc. IEEE International Conference on Image Processing ICIP-97, 1997, p.^pp. 17-20.

14. D. Houlding and J. Vaisey, *Low entropy image pyramids for efficient lossless and progressive image communication*, IEEE Trans. on Image Processing **4** (1995), 1150=1153.

15. B. H. Juang and A. H. Gray, *Multiple stage vector quantization for speech coding*, Proc. Intl. Conf. on Acoust. Speech, and Signal Processing (ICASSP), 1982, p.^pp. 597-600.

16. Y. Kim and W. A. Pearlman, *Lossless volumetric medical image compression*, Proc. SPIE on Applications of Digital Image Processing XXII, 1999, p.^pp. 305-312.

17. H. Lee, Y. Kim, A. H. Rowbey and E. A. Riskin, *Statistical distributions of dct coefficients and their applications to an interframe compression algorithm for 3d medical images*, IEEE Trans. on Med. Imaging **12** (1993), 478-485.

18. Y. Linde, A. Buzo and R. M. Gray, *An algorithm for vector quantizer design*, IEEE Trans. on Commun. **28** (1980), 84-95.

19. J. Pawley (Editor), *Handbook of boilogical confocal microscopy (3rd ed.)*, Plenum, New York, 1995.

20. K. Ramchandran, A. Ortega and M. Vetterli, *Bit allocation for dependent quantization with applications to multiresolution and mpeg video coders*, IEEE Trans. on Image Processing **3** (1994), 533-545.

21. P. Schelkens, A. Munteanu, J. Barbarien, M. Galca, X. Giro-Nieto and J. Cornelis, *Wavelet coding of volumetric medical datasets*, IEEE Trans. on Med. Imaging **22** (2003), no. 3, 441-458.

22. Y. Tao and W. P. Cockshott, *3d microscopic image coding by finite-state vector quantization in an enhanced image pyramid*, SPIE on Medical Imaging 2004, 2004, p.^pp. 1895-1905.

23. D. Taubman and M. W. Marcellin, *Jpeg2000 - image compression: Fundamentals, standards and practice*, Kluwer Academic Publishers, Hingham, MA, 2001.

24. M. Vetterli and C. Herley, *Wavelets and filter banks: Theory and design*, IEEE Trans. on Acoust. Speech Signal Process. **40** (1992), no. 9, 2207-2232.

25. J. Villasenor, B. Belzer and J. Liao, *Wavelet filter evaluation for image compression*, IEEE Trans. on Image Processing **2** (1995), 1053-1060.

26. L. Walras, *Elements of pure economics*, Homewood, Ill: Irwin, 1954.

27. L. Wang and M. Goldberg, *Comparative performance of pyramid data structures for progressive image transmission*, IEEE Trans. on Commun. **39** (1991), 540-548.

28. G. Wolberg, *Digital image warping*, IEEE Computer Society Press, 1994.

29. J. W. Woods and S. D. O'Neill, *Subband coding of images*, IEEE Trans. on Acoust. Speech Signal Process. **34** (1986), 1278-1288.