Chapter 6

Optimal Bit-Allocation for Wavelet Scalable Video Coding with User Preference

Guan-Ju Peng and Wen-Liang Hwang

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/53427

1. Introduction

In this chapter, we introduce the concept and the details of the wavelet-based scalable video coding. The content of this chapter includes the following topics:

- ***Fundamentals of Wavelet-Based Scalable Video Coding :*** The purpose and the general concept of scalable video coding will be introduced in this section. We also give a brief comparison between the major two scalable video coding methods, which are the wavelet-based scalable video coding and H.264/SVC. In addition, we introduce the structure of wavelet-based scalable video coding in this section. A wavelet scalable video encoder consists of two steps. The first step is to decompose each GOP (group of pictures) into multiple subbands, and the second one is to perform entropy coding, which is usually implemented by (embedded zero block coding) [1]. We are going to discuss why and how these two steps achieve video compression and provide universal scalability in a theoretical style.
- ***The Objective Function of Scalable Video Coding :*** The objective of the scalable video coding will be discussed in this section. In the the discussion, the essential elements that affects the performance of scalable video coding will be considered. These essential elements are the status of network transmission, subscribers' preferences, and the video quality in terms of the conventional Peak-to-Noise-Ratio (*PSNR*). And then according to the discussion, an objective function that considers these elements simultaneously is established for the optimization of the scalable video coder.
- ***The Rate Allocation of Wavelet-Based Scalable Video Coding :*** Since the entropy coding procedure needs to know the number of bits allocated (which is usually referred to as *"rate"*) to each subband, the encoder should decides the rates of the subbands before the entropy coding applied to the subbands. In this section, we are going to introduce how to perform rate allocation that optimize the SVC coder with respect to the proposed objective function and compare its performance with those of the existing rate allocation methods. We will also discuss several issues related to the performance of the rate allocation, such as the concept of rate-distortion curve and the inequivalent energy between the pixel and transform domains caused by non-orthogonal filters.

- ***Implementation Issues and Experimental Results :*** We will discuss the computational complexity of the proposed rate allocation method and raises several points that can efficiently reduce the computational time. The experimental results that compare the proposed methods and the existing methods will also be given in the section.
- ***Conclusion and Future Work :*** A conclusive discussion will be given in this section. The discussion will list the contribution of the works mentioned in this chapter and point out some possible issues for the future research.

2. Fundamentals of wavelet-based Scalable Video Coding

Scalable video coding (SVC) encodes a video into a single bitstream comprised of several subset bitstreams. A subset bitstream represents a lower resolution of the video in the spatial, temporal, or quality resolution [2, 3]. SVC is a natural solution to compress the video for a video broadcasting system because the bit-stream generated by the SVC can be divided and separately decoded to support different resolutions. Compared to H.264/SVC [4], which is recently developed based on the prevailing conventional close-loop codec H.264/AVC, the multi-resolution property of 3-D wavelet representation based on motion-compensated temporal filtering (MCTF) is a more natural solution to the scalability issue in video coding [5–8]. However, to compete with the great success of scalable coding methods based on H.264, the MCTF-based 3-D wavelet video codec must be constantly improved.

2.1. The structure of the wavelet scalable video coding

In a MCTF-EZBC wavelet video coding scheme, the video frames are first decomposed into multiple wavelet subbands by spatial and temporal wavelet transforms, then the quantization and entropy coding are sequentially applied to the wavelet subbands. According to the order of the spatial and the temporal decompositions, the wavelet coding schemes can be categorized into two categories : 2D+T (spatial filtering first) and T+2D (temporal filtering first). However, no matter 2D+T or T+2D scheme is applied, the spatial and the temporal filterings can be described independently.

The purpose of spatial filtering is separating low and high frequency coefficients from a video frame. The spatial filtering usually consists of multiple sequential 2D wavelet decompositions. In a 2D wavelet decomposition, the input signal, which is represented by a N by N two dimensional matrix, is decomposed into four $N/2$ by $N/2$ two dimensional matrices, which are denoted by LL, HL, LH, and HH. For these subbands, the previous letter means that the subband contains the low (L) or high (H) frequency coefficients after the horizontal 1D wavelet transform and the following letter means the subband contains the low (L) or high (H) frequency coefficients after the vertical 1D wavelet transform. After the decomposition, subband LL is sequentially taken as the input of the next level spatial decomposition.

If we let $\mathcal{H}_{k,0}$ and $\mathcal{H}_{k,1}$ denote the analyzing matrices in the k-th spatial decomposition, the corresponding wavelet subbands can be computed as

$$\begin{bmatrix} F_{k0} & F_{k1} \\ F_{k2} & F_{k3} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_{k0} \\ \mathcal{H}_{k1} \end{bmatrix} F_{(k-1)0} \begin{bmatrix} \mathcal{H}_{k0}^T & \mathcal{H}_{k1}^T \end{bmatrix} \quad (1)$$

http://dx.doi.org/10.5772/53427

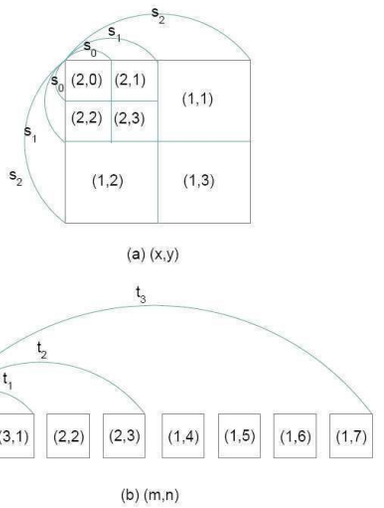

Figure 1. An example of indexing the spatial and temporal decomposed subbands. The above figure shows the indices of the spatially decomposed subbands, in which the subband indexed by (x, y) represents the subband is the y-th after the x-th spatial decomposition. The below figure shows the indices of the temporally decomposed subbands, in which the subband indexed by (m, n) represents the subband is the n-th after the m-th spatial decomposition. Accordingly, any subband obtained by the MCTF can be indexed by (xy, mn).

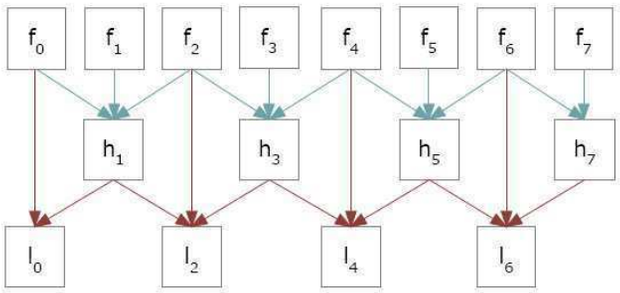

Figure 2. An example of the lifting structure used in the temporal decomposition. The subbands h, which contain high frequency coefficients, are first obtained after the prediction stage. Then the subbands l, which contain the low frequency coefficients, are obtained after the update stage.

where F_{k0}, F_{k1}, F_{k2}, and F_{k3} correspond to the LL, HL, LH, HH subbands, respectively. If there is another spatial decomposition applied to the frame, the subband F_{k0} is further decomposed by the $k+1$-th analyzing matrices $\mathcal{H}_{k+1,0}$ and $\mathcal{H}_{k+1,1}$. According to this definition, the subband F_{00} is the frame before any spatial decomposition and any wavelet subband after the spatial filtering can be indexed by xy, which represent the subband is the y-th subband after the x-th spatial decomposition. In Figure 1(a), the example shows the wavelet subbands with the indices after performing the spatial filtering consisting of two spatial decompositions.

To reconstruct $F_{(k-1)0}$ from the wavelet subbands F_{k0}, F_{k1}, F_{k2}, and F_{k3}, the synthetic matrices $\mathcal{G}_{k0}$ and $\mathcal{G}_{k1}$ are used in the synthetic procedure, which can be represented by several matrix operations as

$$F_{(k-1)0} = \begin{bmatrix} \mathcal{G}_{k0} & \mathcal{G}_{k1} \end{bmatrix} \begin{bmatrix} F_{k0} & F_{k1} \\ F_{k2} & F_{k3} \end{bmatrix} \begin{bmatrix} \mathcal{G}_{k0}^T \\ \mathcal{G}_{k1}^T \end{bmatrix}. \tag{2}$$

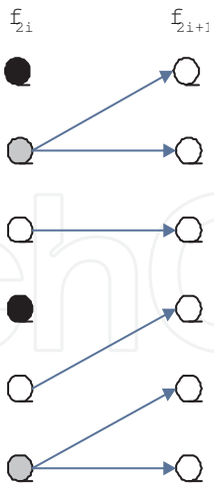

Figure 3. An example of MCTF motion estimation. The types of connectivity pixels in the example are single-connected, multiple-connected, and unconnected pixels. The corresponding prediction (P) and update (U) matrices are given in Eq.(5), where $P^{2i,2i+1}[x,y] = 1$ indicates that the x-th pixel in frame f_{2i+1} is predicted by the y-th pixel in frame f_{2i}. Note that $U^{2i+1,2i}(x,y) = 1$ means the x-th pixel in frame f_{2i} is updated by the y-th pixel in frame f_{2i+1}.

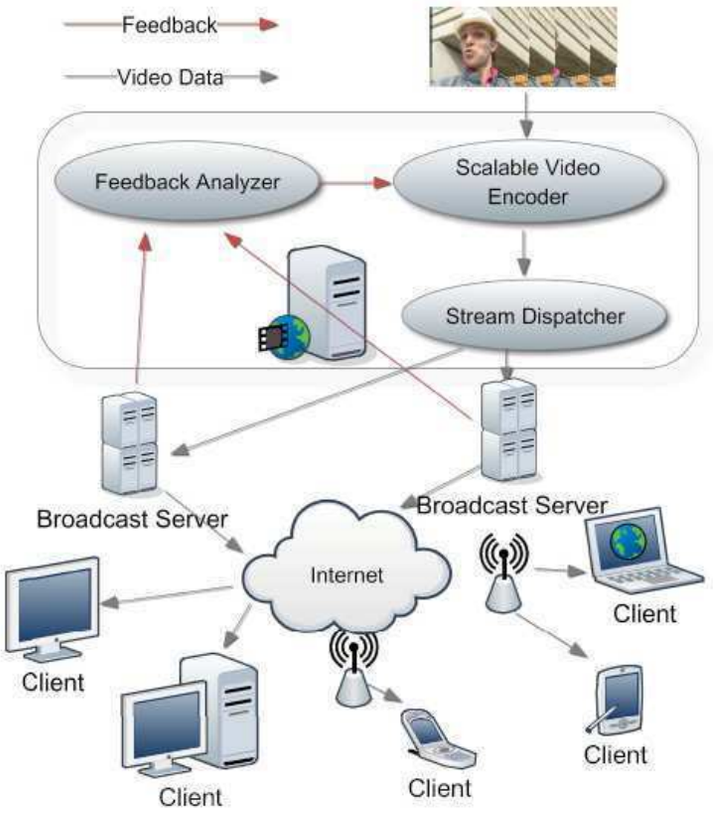

Figure 4. A video broadcasting system usually contains a video source, a scalable video coder, broadcasting servers, the network, and subscribers. Users' information can be delivered to the servers via feedback links.

http://dx.doi.org/10.5772/53427

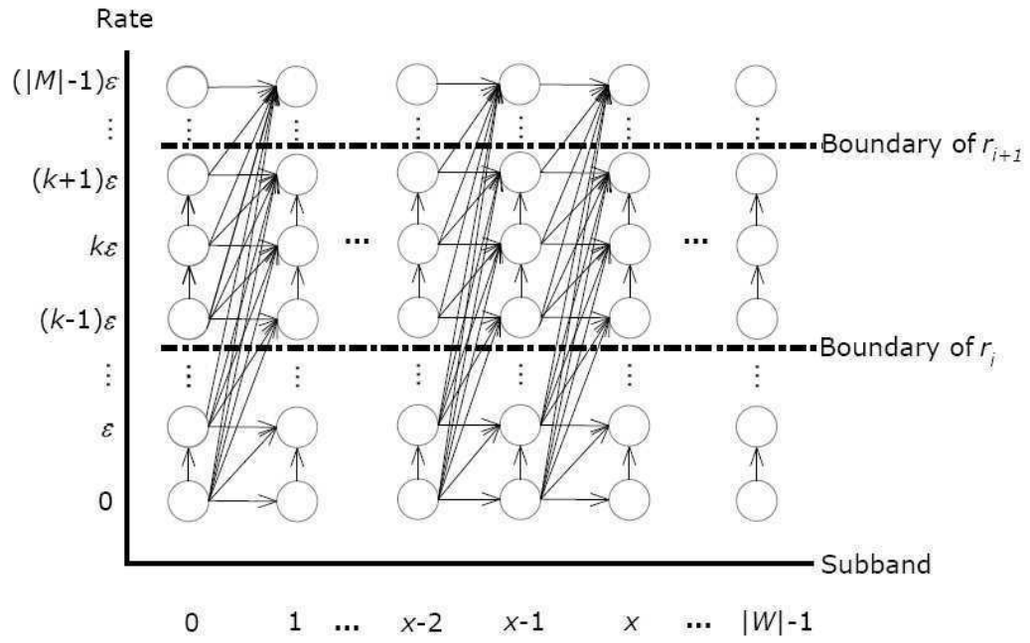

Figure 5. The construction of the DP graph. The x-axis indicates the subbands and the y-axis indicates the rate. For an arc, we calculate the number of bits assigned to the subband associated with the sink node of the arc, and derive the weighted distortion of the subband. For any node on a path, we can calculate the number of bits assigned to the subband corresponding to the node as well as its weighted distortion. The node (x,k) maps to subband $seq^{-1}(x)$ and to quality resolution $q(k) = r_{i+1}$. Note that all the nodes indexed by $(.,k-1)$, $(.,k)$, and $(.,k+1)$ belong to quality resolution r_{i+1}.

Compared to the spatial filtering, which takes a frame as its input, the temporal filtering takes multiple frames (T+2D) or the subands with the same spatial indices (2D+T) as its input. The temporal filtering consists of multiple temporal decompositions. However, unlike the spatial filtering, which only needs several 2-D wavelet transforms during the procedure, the temporal filtering needs multiple 1-D wavelet transforms and motion estimation/compensation to specify the input coefficients of each wavelet transform. So the temporal filtering is more complicated compared to the spatial filtering.

Usually, a temporal decomposition is implemented by the lifting structure, which consists of the predicting stage and the updating stage. As depicted in Figure 2, the H-frames (denoted by h) are firstly obtained after the predicting stage, then the L-frames (denoted by l) are obtained after the updating stage. The lifting structure can be described by the matrix operations, in which a frame is represented by a one dimensional matrix f. If each input frame of the temporal filtering has the size N by N and can be represented by a N by N two dimensional matrix F, the one dimensional matrix f has the size N^2 and can be mapped from F by letting $f[i*N+j] = F[i,j]$.

The motion vectors are computed before the predicting and updating stages. We use the $P_m^{x,y}$, which is a two dimensional N^2 by N^2 matrix, to denote the motion vectors obtained by predicting the y-th frame from the x-th frame in the m-th temporal decomposition. Then the predicting stage of the m-th temporal decomposition can be written as

$$h_m^{2i+1} = f_{m-1}^{2i+1} - (\mathcal{H}_m[2i]P_m^{2i,2i+1}f_{m-1}^{2i} + \mathcal{H}_m[2i+2]P_m^{2i+2,2i+1}f_{m-1}^{2i+2}), \tag{3}$$

where $\mathcal{H}_m$ is the corresponding coefficient of the wavelet filter. Since the index m represents the m-th lifting stage, the term f_{m-1}^{i} represents the i-th frame which is obtained after the $(m-1)$-th lifting stage. If the 5-3 wavelet transform is used, the value of $\mathcal{H}_m$ in the predicting stage is -0.5.

In the updating stage, the inverse motion vector matrix $U_m^{y,x}$, which has the same size as $P_m^{x,y}$ and can be calculated from the matrix $P_m^{x,y}$ [9], is used to compute the decomposed L-frame as

$$l_m^{2i} = f_{m-1}^{2i} + (\mathcal{H}_m[2i+1]U_m^{2i+1,2i}h_m^{2i+1} + \mathcal{H}_m[2i-1]U_m^{2i-1,2i}h_m^{2i-1}), \tag{4}$$

where $\mathcal{H}_m$ is the corresponding coefficient of the wavelet filter. If the 5-3 wavelet transform is used, the value of $\mathcal{H}_m$ in the updating stage is 1.

To understand how to represent the motion vectors with a matrix, an example is given in Figure 3. And these two matrices are constructed as follows:

$$P^{2i,2i+1} = \begin{bmatrix} 0&1&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \\ 0&0&0&0&0&1 \end{bmatrix}, \quad U^{2i+1,2i} = \begin{bmatrix} 0&0&0&0&0&0 \\ 1&0&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \end{bmatrix}. \tag{5}$$

After the m-th temporal decomposition, if another temporal decomposition is needed for the temporal filtering, the L-frames generated by the current temporal decomposition are taken as its input. So we let f_m^i, which is the i-th frame of the $(m+1)$-th temporal decomposition, as l_m^{2i}, which is the output $2i$-th frame of the m-th temporal decomposition. Although the H-frames do not participate in the $(m+1)$-th temporal decomposition, we still arrange the indices to them by letting $f_m^{i+S} = h_m^{2i+1}$, in which S is the number of the input frames in the $(m+1)$-th temporal decomposition. So any frames obtained by the temporal filtering can be indexed by mn, which represents it is the n-th subband after the m-th temporal decomposition.

Usually, the temporal decompositions are sequentially performed until the output has only one L-frame. Since the decomposed frame can be synthesized, only the frames which can not be synthesized are necessary to reconstruct all the frames. In Figure 1(b), an example, in which the size of the group of pictures (GOP) is 8, shows the indices of the frames that can not be synthesized from the decomposed frames.

To recover the original frames, the synthesis is applied to the decomposed frames. The frame which is decomposed last is synthesized first, and vice versa. In the procedure of the synthesis, the inverse updating is firstly performed as

$$f_{m-1}^{2i} = l_m^{2i} - (\mathcal{G}_m[2j+1]U_m^{2j+1,2i}h_m^{2j+1} + \mathcal{G}_m[2j-1]U_m^{2j-1,2i}h_m^{2j-1}), \tag{6}$$

http://dx.doi.org/10.5772/53427

where $\mathcal{G}_m$ is the coefficient of the wavelet transform used in the inverse updating stage. If the temporal 5-3 tap filter is used, the value of $\mathcal{G}_m$ is -1 in the inverse updating stage. After the inverse updating stage, the inverse predicting stage is performed as

$$f_{m-1}^{2i+1} = h_m^{2i+1} + (\mathcal{G}_m[2j]P_m^{2j,2i+1} f_{m-1}^{2j} + \mathcal{G}_m[2j+2]P_m^{2j+2,2i+1} f_{m-1}^{2j+2}). \quad (7)$$

If the temporal 5-3 filter is used, the value of $\mathcal{G}_m$ is 0.5 in the inverse updating stage. For some wavelet filters, such as $9-7$ filter, a temporal decomposition needs multiple level lifting structure, and can be easily extended by cascading multiple one level lifting structures.

After the spatial and temporal filterings, the quantization and entropy coding are applied to these wavelet subbands. The coefficients of these subbands may be quantized by scalar or vector quantization, then the quantized coefficients are coded without loss by the entropy coder. The method is common in the still image compression standard, such as JPEG 2000 [10]. In the decoding process, the quantized coefficients are obtained by decoding the received bitstream, then the subbands are rescaled according to the quantization step used in the encoder. The advantage of separating the quantization and the entropy coding is that the quality of the reconstructed video can be predicted according to the quantization step.

The quantization and the entropy coding can also be combined with the bitplane coding method, such as the EZBC entropy coder [1]. In these methods, the rates allocated to the subbands are calculated first, then the entropy coder encodes the subbands with these rates. The advantage of the scheme is the rates of the subbands can be any non-negative integers and the performance of the bit-allocation can be improved accordingly.

3. The objective function of Scalable Video Coding

Many researchers employ a weighting coefficient to represent the relative importance of a resolution. For example, Ramchandran, Ortega, and Vetterli [11] model the distortion as the summation of the weighted mean-square-errors (MSE) on different resolutions, and propose a bit-allocation algorithm based on the exhaustive search technique. Schwarz and Wiegrand [12] adopt a similar approach by weighting the MSEs of the base layer and the enhancement layer, and demonstrate the effect of employing different weightings on each layer on the overall coding performance. The above works do not explain the meaning of the weights or how to derive them. Because the peak-signal-to-noise ratio (PSNR) is most commonly used as a quality measurement of a coding system, instead of weighting the MSE of a resolution, we weight the PSNR as a measurement of relative importance of the resolution to the overall coding performance.

3.1. Design the objective function of Scalable Video Coding

A good coding performance metric for SVC should consider the subscriber's preference for different resolutions. For example, if we want to produce bitstreams in two scenarios: one where all the subscribers prefer the QCIF display and the other where all the subscribers prefer the CIF display, then the optimal bitstreams for the two scenarios should be different. In the first scenario, the optimal bit-allocation can only be obtained by allocating all the bits to the subbands that support the QCIF display. Obviously, this allocation cannot be optimal for

the second scenario in which the optimal bit-allocation must encode more spatial subbands to support a higher spatial resolution display with the CIF format.

A general video broadcasting system consists of a video source, a scalable video coder, broadcasting servers, the network, and subscribers, as shown in Figure 4. The scalable coder encodes a source video so that the network's bandwidth requirement can be met and the subscriber's demand can be satisfied. The satisfaction of the subscriber's demand can be quantified to measure the system's performance. In [3], the performance of SVC is measured as follows:

$$Q_{all} = \sum_{i \in N} Q_i, \tag{8}$$

where N denotes the set of subscribers, and Q_i denotes the satisfaction of subscriber i's demand by SVC, which is usually measured by the $PSNR$. However, we found that the $PSNR$ is not sufficient to satisfy the demand of a subscriber because he/she may prefer higher frame rates or spatial resolutions than the $PSNR$. Thus, we introduce the preference factor $\psi \in [0,1]$ for each subscriber and combine it with the $PSNR$ to obtain the following performance measurement:

$$Q_{all} = \sum_{i \in N} \psi_i PSNR_i. \tag{9}$$

If we let **S**, **T**, and **R** denote the sets of spatial, temporal, and quality resolutions respectively, then a resolution in SVC can be represented by (s,t,r), where $s \in \mathbf{S}$, $t \in \mathbf{T}$, and $r \in \mathbf{R}$. Denote subscriber i's preference for the resolution (s,t,r) as $\psi_{i,(s,t,r)}$, and let the $PSNR$ of the resolution be $PSNR_{(s,t,r)}$. Then, Equation (9) can be re-written as follows:

$$Q_{all} = \sum_{s \in \mathbf{S}, t \in \mathbf{T}, r \in \mathbf{R}} PSNR_{(s,t,r)} \sum_{i \in N} \psi_{i,(s,t,r)}. \tag{10}$$

The performance measurement can be normalized based on the subscriber's preference so that we obtain

$$\begin{aligned} Q_{average} &= \frac{Q_{all}}{\sum_{s \in \mathbf{S}, t \in \mathbf{T}, t \in \mathbf{R}} \sum_{i \in N} \psi_{i,(s,t,r)}} \\ &= \sum_{s \in \mathbf{S}, t \in \mathbf{T}, r \in \mathbf{R}} PSNR_{(s,t,r)} \left(\frac{\sum_{i \in N} \psi_{i,(s,t,r)}}{\sum_{s \in \mathbf{S}, t \in \mathbf{T}, t \in \mathbf{R}} \sum_{i \in N} \psi_{i,(s,t,r)}} \right) \\ &= \sum_{s \in \mathbf{S}, t \in \mathbf{T}, r \in \mathbf{R}} PSNR_{(s,t,r)} \mu_{(s,t,r)}. \end{aligned} \tag{11}$$

Because $\mu_{(s,t,r)}$ considers the preferences of all subscribers for the resolution (s,t,r), it can be regarded as the preference of the system to the resolution. Moreover, from the definition of $\mu_{(s,t,r)}$, we have $\mu_{(s,t,r)} \geq 0$ and

$$\sum_{s \in \mathbf{S}, t \in \mathbf{T}, r \in \mathbf{R}} \mu_{(s,t,r)} = 1. \tag{12}$$

http://dx.doi.org/10.5772/53427

The $PSNR_{(s,t,r)}$ can be calculated as follows:

$$PSNR_{(s,t,r)} = 10log_{10}\frac{255^2}{\bar{D}_{(s,t,r)}}, \tag{13}$$

where $\bar{D}_{(s,t,r)}$ denotes the average mean square error (MSE) of the frames in resolution (s,t,r). If we substitute Equation (13) into Equation (11) and use Equation (12), we have

$$\begin{aligned} Q_{average} &= 10\log_{10}255^2 \sum_{s\in\mathbf{S},t\in\mathbf{T},r\in\mathbf{R}} \mu_{(s,t,r)} - \sum_{s\in\mathbf{S},t\in\mathbf{T},r\in\mathbf{R}} \mu_{(s,t,r)}\log_{10}\bar{D}_{(s,t,r)} \\ &= 10\log_{10}255^2 - log_{10}(\prod_{s\in\mathbf{S},t\in\mathbf{T},r\in\mathbf{R}} \bar{D}_{(s,t,r)}^{\mu_{(s,t,r)}}). \end{aligned} \tag{14}$$

It is obvious that maximizing the average performance $Q_{average}$ is equivalent to minimizing the geometric mean of the distortion

$$\prod_{s\in\mathbf{S},t\in\mathbf{T},r\in\mathbf{R}} \bar{D}_{(s,t,r)}^{\mu_{(s,t,r)}}. \tag{15}$$

Note that, in SVC, each temporal resolution involves a different number of frames. If a scalable coder adopts the dyadic temporal structure, which assumes that the number of frames in temporal resolution t is 2^t, then the overall distortion of the resolution (s,t,r) in a GOP is

$$D_{(s,t,r)}^{GOP} = 2^t\bar{D}_{(s,t,r)}. \tag{16}$$

3.2. Subband Weighting

The weighting factor indicates how much a unit quantization power in the subband contributes to the overall distortion in the reconstructed GOP. That is, for the subband indexed by z, its weighting $\gamma(z)$ is given to satisfy

$$D_{(s,t,r)}^{GOP} = \sum_{z\in W(s,t)} \gamma(z)\times D_z \tag{17}$$

where $D_{(s,t,r)}^{GOP}$ is the variance of the distortion in the pixel domain, D_z is the variance of subband's distortion in the wavelet domain, and $W(s,t)$ is the set including the subbands necessary to reconstruct the resolution (s,t,r). The weighting factor $\gamma(z)$ can be computed by

$$\gamma(z) = \alpha(z)\times\beta(z) \tag{18}$$

where $\alpha(z)$ is the spatial weighting factor and $\beta(z)$ is the temporal weighting factor.

The spatial weighting factors can be directly computed according to the error propagation model mentioned in [13]. However, to compute the temporal weighting factors, the effect of the motion compensation must also be considered, and the approach is mentioned in [14]. Since the computational complexity is extremely high, a fast algorithm is proposed in [15] to increase the computing speed.

3.3. Formulation of the rate-distortion function

In this sub-section, we formulate the rate-distortion function of a wavelet-based scalable video coder. We use non-negative integers to index the spatial and temporal resolutions. The lowest resolution is indexed by 0, and a higher resolution is indexed by a larger number. Let $\mathbf{p}$ and $\mathbf{q}$ denote the number of spatial and temporal decompositions respectively; then, the spatial resolution index s and temporal resolution index t are in the ranges $\{0, 1, \cdots, \mathbf{p}\}$ and $\{0, 1, \cdots, \mathbf{q}\}$ respectively. Note that we use (xy, mn) to denote the spatial-temporal subband, which is the y-th spatial subband after the x-th spatial decomposition and the n-th temporal subband after the m-th temporal decomposition. Thus, if we let $W_{s,t}$ denote the set of subbands used to reconstruct the video of spatial resolution s and temporal resolution t, then, $W_{s,t}$ is comprised of

$$\begin{aligned}\{(xy, mn) | x = \mathbf{p} - s + 1, ..., \mathbf{p}; y = 1, 2, 3; m = \mathbf{q} - t + 1, ..., \mathbf{q}; n = 2^{\mathbf{q}-m}, ..., 2^{\mathbf{q}-m+1} - 1\} \cup \\ \{(\mathbf{p}0, mn) | m = \mathbf{q} - t + 1, \cdots, \mathbf{q}; n = 2^{\mathbf{q}-m}, \cdots, 2^{\mathbf{q}-m+1} - 1\} \cup \\ \{(xy, \mathbf{q}0) | x = \mathbf{p} - s + 1, ..., \mathbf{p}, y = 1, 2, 3\} \cup \{(\mathbf{p}0, \mathbf{q}0)\}. \quad (19)\end{aligned}$$

Figure 1 shows an example of two spatial and three temporal resolutions.

We also assume that all the subscribers receive the same bitstream containing the quality resolution r; therefore, a subscriber to resolution (s, t, r) can decode the substream corresponding to the subbands that support the spatial resolution s and the temporal resolution t. For each quality resolution r, let $\beta_r[(xy, mn)]$ represent the number of bits assigned to subband (xy, mn) for the quality resolution. This assumption simplifies our bit-allocation analysis significantly because we only need to consider the distribution of the bits for the quality resolutions. Obviously, we have

$$\beta_{r+1}[(xy, mn)] \geq \beta_r[(xy, mn)] \quad (20)$$

for each subband (xy, mn). Let b_r denote the maximum number of bits for all the subbands of quality resolution r in a GOP, and let W be the set of all subbands; then, the bit constraint for the quality resolution r can be written as

$$b_r = \sum_{z \in W} \beta_r[z], \quad (21)$$

where z ranges over all subbands in W. Recall that the average distortion of all the subbands of the frames in the resolution (s, t, r) is represented by $\bar{D}^w_{(s,t,r)}$. We introduce a new notation

http://dx.doi.org/10.5772/53427

$\Theta(s,t,\beta_r)$ for $\bar{D}^w_{(s,t,r)}$ to explicitly represent the average distortion in the wavelet domain as a function of B_r. According to Equation (17), we have

$$\Theta(s,t,\beta_r) = \frac{1}{2^t} \sum_{z \in W_{s,t}} w_z^{s,t} \bar{D}_z^w(\beta_r[z])], \tag{22}$$

where $\bar{D}_z^w(\beta_r[z])$ indicates the average distortion of subband z encoded with $\beta_r[z]$ bits. Substituting the subbands support for the resolution (s,t,r), defined in Equation (19), for $W_{s,t}$ in Equation (22), we obtain

$$\begin{aligned}\Theta(s,t,\beta_r) = \frac{1}{2^t}\Big(& \sum_{m=\mathbf{q}-t+1}^{\mathbf{q}} \sum_{n=2^{\mathbf{q}-m}}^{2^{\mathbf{q}-m+1}-1} \sum_{x=\mathbf{p}-s+1}^{\mathbf{p}} \sum_{y=1}^{3} w^{s,t}_{(xy,mn)} \bar{D}^w_{(xy,mn)}(\beta_r[(xy,mn)]) \\ & + \sum_{m=\mathbf{q}-t+1}^{\mathbf{q}} \sum_{n=2^{\mathbf{q}-m}}^{2^{\mathbf{q}-m+1}-1} w^{s,t}_{(\mathbf{p}0,mn)} \bar{D}^w_{(\mathbf{p}0,mn)}(\beta_r[(\mathbf{p}0,mn)]) \\ & + \sum_{x=\mathbf{p}-s+1}^{\mathbf{p}} \sum_{y=1,2,3} w^{s,t}_{(xy,\mathbf{q}0)} \bar{D}^w_{(xy,\mathbf{q}0)}(\beta_r[(xy,\mathbf{q}0)]) \\ & + w^{s,t}_{(\mathbf{p}0,\mathbf{q}0)} \bar{D}^w_{(\mathbf{p}0,\mathbf{q}0)}(\beta_r[(\mathbf{p}0,\mathbf{q}0)])\Big). \end{aligned} \tag{23}$$

Let $\mathbf{v}$ be the number of quality resolutions indexed from $0 \cdots, \mathbf{v}-1$, and let $\{\beta_0, \cdots, \beta_{\mathbf{v}-1}\}$ be the bit-allocation profile. We can represent Equation (15) explicitly as $\mathcal{D}(\beta_0, \cdots, \beta_{\mathbf{v}-1})$ to indicate the dependence of the average distortion of a GOP on the bit-allocation profile. Then, we obtain

$$\mathcal{D}(\beta_0, \cdots, \beta_{\mathbf{v}-1}) = \prod_{r=0}^{\mathbf{v}-1} \prod_{s=0}^{\mathbf{p}-1} \prod_{t=0}^{\mathbf{q}-1} \bar{D}^{\mu_{(s,t,r)}}_{(s,t,r)} \tag{24}$$

$$= \prod_{r=0}^{\mathbf{v}-1} \prod_{s=0}^{\mathbf{p}-1} \prod_{t=0}^{\mathbf{q}-1} \Theta(s,t,\beta_r)^{\mu_{(s,t,r)}} \tag{25}$$

$$= \prod_{r=0}^{\mathbf{v}-1} \mathcal{D}_r(\beta_r), \tag{26}$$

where $\mathcal{D}_r(\beta_r)$ is the average distortion of quality resolution r when the subscriber's preference factor $\mu_{(s,t,r)}$ is considered in weighting $\Theta(s,t,\beta_r)$.

The rate-distortion problem (P) can now be formulated as finding the bit-allocation profile $\{\beta_0, \cdots, \beta_{\mathbf{v}-1}\}$ that satisfies the constraints in Equations (20) and (21) and minimizes the distortion function specified in Equation (26):

$$\begin{gathered} \min \mathcal{D}(\beta_0, \beta_1, \cdots, \beta_{\mathbf{v}-1}) \\ \text{subject to } \sum_{z \in W} \beta_i(z) = b_i, \text{ for } i = 0, \cdots, \mathbf{v}-1, \\ \text{and } \beta_{i-1}(z) \le \beta_i(z) \text{ for } i = 1, \cdots, \mathbf{v}-1, \end{gathered} \tag{27}$$

4. The rate allocation of wavelet-based Scalable Video Coding

The optimal bit-allocation problem (P) can be solved by solving a sequence of bit-allocation sub-problems (P_r), with quality resolution $r = 0, \cdots, \mathbf{v}-1$. The sub-problem (P_r) is defined as follows:

$$\begin{aligned} &\min \mathcal{D}(\beta_0, \beta_1, \cdots, \beta_{r-1}, \beta_r, \beta_r, \cdots, \beta_r) \\ &\text{subject to } \textstyle\sum_{z\in W} \beta_i(z) = b_i, \text{ for } i = 0, \cdots, r, \\ &\text{and } \beta_{i-1}(z) \le \beta_i(z) \text{ for } i = 1, \cdots, r, \end{aligned} \tag{28}$$

where W is the set of all subbands, and $\{b_i\}$ is a given non-decreasing sequence that corresponds to the bit constraints. The problem (P_r) allocates bits from the quality resolution 1 to r; hence, all the subscriptions for a quality resolution $> r$ will use the bit-allocation result of the quality resolution r. Thus, we have $\beta_i = \beta_r$ for $i = r+1, \cdots, \mathbf{v}-1$ in Equation (28). The optimal bit-allocation problem (P) can be solved by solving (P_0), followed by (P_1) based on the solution of (P_0), and so on up to solving $(P_{\mathbf{v}-1})$. In the following subsections, we propose two methods to solve (P_r). The first finds the upper bound of (P_r) by a Lagrangian-based approach, and the second finds the exact solution by using the less efficient dynamic programming approach.

4.1. Lagrangian-based solution

The bit-allocation problem is usually analyzed by the Lagrangian multiplier method. By assuming that $\prod_{i=a}^{b} f_i = 1$ for any function f_i with $b > a$, the objective of (P_r) can be re-written as

$$\begin{aligned} &\mathcal{D}(\beta_0, ..., \beta_{r-1}, \beta_r, ..., \beta_r) \\ &= \prod_{s=0}^{\mathbf{p}-1} \prod_{t=0}^{\mathbf{q}-1} \{ \prod_{k=0}^{r-1} \Theta(s,t,\beta_k)^{\mu_{(s,t,k)}} \prod_{k=r}^{\mathbf{v}-1} \Theta(s,t,\beta_r)^{\mu_{(s,t,k)}} \}. \end{aligned} \tag{29}$$

Because $\sum_{s\in\mathbf{S},t\in\mathbf{T},r\in\mathbf{R}} \mu_{(s,t,r)} = 1$ (See Equation (12)), by applying the generalized geometric mean - arithmetic mean inequality to Equation (29), we can obtain its upper bound as follows:

$$\mathcal{D}(\beta_0, ..., \beta_{r-1}, \beta_r, ..., \beta_r) \le C + \sum_{s=0}^{\mathbf{p}-1} \sum_{t=0}^{\mathbf{q}-1} \sum_{k=r}^{\mathbf{v}-1} \mu_{(s,t,k)} \Theta(s,t,\beta_r), \tag{30}$$

where the constant $C = \sum_{s=0}^{\mathbf{p}-1} \sum_{t=0}^{\mathbf{q}-1} \sum_{k=0}^{r-1} \mu_{(s,t,k)} \Theta(s,t,\beta_k)$. Note that the constant C is computed from the bit-allocation results from resolution 0 to resolution $r-1$.

Now we can find the solution for the problem (P_r^+), which is the upper bound of the problem (P_r). The problem (P_r^+) is defined as

$$\begin{aligned} &\min_{\beta_r} \Omega_r(\beta_r) = \textstyle\sum_{s=0}^{\mathbf{p}-1} \sum_{t=0}^{\mathbf{q}-1} \sum_{k=r}^{\mathbf{v}-1} \mu_{(s,t,k)} \Theta(s,t,\beta_r) \\ &\text{subject to } \textstyle\sum_{z\in W} \beta_r(z) = b_r. \end{aligned} \tag{31}$$

http://dx.doi.org/10.5772/53427

After substituting Equation (22) into Equation (31) for $\Theta(s,t,\beta_r)$ and re-arranging the terms, we have

$$\Omega_r(\beta_r) = \sum_{z \in W} \rho_z^r \bar{D}_z(\beta_r[z]), \tag{32}$$

where z is a subband, W is the set of all subbands, and ρ is the final weighting factor. Note that ρ is computed from the preference weighting μ and the spatial-temporal weighting $w^{s,t}$.

Now, replacing $\Omega_r(\beta_r)$ in Equation (31) with Equation (32), the problem (P_r^+) can be solved optimally by using the Lagrangian approach with the Lagrangian function:

$$L(\lambda, \beta_r) = \sum_{z \in W} \rho_z^r \bar{D}_z(\beta_r[z]) - \lambda(b_r - \sum_{z \in W} \beta_r[z]). \tag{33}$$

A necessary condition for optimal bit-allocation can be satisfied by taking the partial derivative with respect to λ and β_r and setting the results to zero. Thus, the optimal bit-allocation vector β_r^* for the quality resolution r must satisfy

$$\frac{\partial L(\lambda, \beta_r)}{\partial \beta_r[z]} = \rho_z^r \frac{\partial \bar{D}_z(\beta_r[z])}{\partial \beta_r[z]} + \lambda = 0, \tag{34}$$

and

$$\frac{\partial L(\lambda, \beta_r)}{\partial \beta_r[z]} = b_r - \sum_{z \in W} \beta_r[r] = 0. \tag{35}$$

The two necessary conditions require that 1) the optimal bit-allocation β_r^* must exist when the rate-distortion functions of all the subbands have the same weighted slope; and 2) at that particular slope, the total number of bits of all the subbands is b_r.

It is straightforward to show that if $\bar{D}_z(\beta_r[z])$ is convex for any z, then $\sum_{z \in W} \rho_z^r \bar{D}_z(\beta_r[z])$ with $\rho_z^r \geq 0$ is a convex function; therefore, the necessary condition is also the sufficient condition for β_r^*. We can use a similar approach to that in [16] to derive an efficient algorithm to find the optimal bit-allocation vector. Thus, we modify the distortion function $\bar{D}_z(\beta_r[z])$ to make it a convex function. We initialize $\beta_r[z] = 0$ for all subbands, and divide b_r into $\lceil b_r/\delta \rceil$ segments with δ bits for each segment. In each stage of our algorithm, we calculate $\bar{D}_z(\beta[z] + \delta)$ and select the subband z' that has the largest weighted absolute slope:

$$\arg\max_{z \in W} \rho_z^r \frac{|\bar{D}_z(\beta[z] + \delta) - \bar{D}_z(\beta[z])|}{\delta}. \tag{36}$$

Then, we only modify the bit-allocation vector of the component that corresponds to the subband z' by letting

$$\beta_r[z'] \leftarrow \begin{cases} \beta_r[z'] + \delta, \text{ if } \sum_{z \in W} \beta_r[z] \leq b_r - \delta, \\ \beta_r[z'] + (b_r - \sum_{z \in W} \beta_r[z]), \text{ otherwise .} \end{cases} \tag{37}$$

We repeat the above process several times until the constraint is achieved.

4.2. Optimal solution based on dynamic programming

Although the proposed Lagrangian-based method is efficient and theoretically sound, it optimizes the upper bound of the true objective function. In this section, we propose another optimal bit-allocation method based on dynamic programming (DP). Although the proposed method uses more memory and requires more computation time than the Lagrangian-based method, it can find the optimal bit-allocation for the true objective function.

To solve the bit-allocation problem with the DP-based method, we represent the problem as an acyclic directed graph $G = (N, A)$, called a DP graph for short, where N is the set of nodes and the members of A are arcs. The arc from node i_k to node i_l is represented by $i_k \rightarrow i_l$ where i_k and i_l are the source node and the sink node of the arc respectively. A path can be represented as a concatenation of arcs.

Let seq be a bijection mapping from the subbands to the integer set from 0 to $|W| - 1$, where $|W|$ is the number of subbands; and let seq^{-1} be its inverse mapping from an integer to a subband. To construct the DP graph for the problem (P), we arrange the subbands z as a sequence $seq(z) \in \{0, \cdots, |W| - 1\}$, and divide $b_{\mathbf{v}-1}$ into $M = \lceil \frac{b_{\mathbf{v}-1}}{\epsilon} \rceil$ components. For convenience, we set the bit constraint b_r with $r = 0, \cdots, \mathbf{v} - 2$ as an integer multiple of ϵ. Then, we introduce the function q, which maps a node in the DP graph to its corresponding quality resolution.

The nodes in the DP graph are $\{(seq(z), k) |\ k = 0, 1, \cdots, M, z \in W\}$. If we let (a, b) be the source node of the arc $(a, b) \rightarrow (seq(z), k)$ to the sink node $(seq(z), k)$, then depending on the position of $(seq(z), k)$, (a, b) belongs to the set $\{(seq(z) - 1, i) | i \leq k\} \cup \{(seq(z), k - 1)\}$, $\{(0, k - 1)\}$, or $\{(seq(z) - 1, 0)\}$.

Figure 5 shows the constructed DP graph. For the node (x, k), the corresponding subband is $seq^{-1}(x)$ and the corresponding quality resolution is $q(k) = r_{i+1}$. We can calculate the number of bits assigned to the subband associated with the sink node of each arc in the DP graph as well as the weighted distortion of the subband. For example, if the arc is $(u, k_u) \rightarrow (v, k_v)$, then the number of bits assigned to the subband $seq^{-1}(v)$ is $(k_v - k_u)\epsilon$. Accordingly, we can also calculate the number of bits assigned to each subband on any path, which consists of consecutive arcs, in the DP graph.

The DP approach uses $|W|$ passes to solve the problem (P) with the constraint that at the end of pass i, the optimal path to each node in the DP graph indexed by $(i, .)$ is found and recorded. At the first pass 0, we find the number of bits allocated to any node $(0, j)$ with $j \in \{0, \cdots, M - 1\}$. Then, based on that result, at pass 1, we find and record the optimal bit-allocation path (the path with the smallest weighted distortion among all the paths that end at the node) to each node $(1, .)$. Based on the result of the previous pass $i - 1$, we can repeat the process to derive the optimal bit-allocation path to any node $(i, .)$ and record the path to the node. After the pass $|W| - 1$, the bit-allocation corresponding to the optimal path from $(0, 0)$ to $(|W| - 1, k_l)$ with $k_l\epsilon = b_r$ for some b_r is the optimal bit-allocation to the quality resolution r; i.e., the optimal solution for problem (P_r) is derived. It can be shown that the DP approach ensures there is only one optimal path beginning at $(0, 0)$ to any node in the DP graph, but the proof is omitted due to the space limitation.

http://dx.doi.org/10.5772/53427

4.3. Min-max approach for unknown preferences

The optimization algorithms presented in previous sections require the subscriber's preference information. However, in many applications, the preference is not available to the encoder. Thus, we present a min-max approach that finds the optimal bit-allocation when the subscriber's preference is not known. The approach is a conservative strategy that ensures the worst performance of the algorithm is above a certain quality.

Let $\mu = [\mu_{(s,t,r)}]$ denote the subscriber's preference vector. In addition, let $\beta = [\beta_0, \cdots, \beta_{\mathbf{v}-1}]$ and $b = [b_0, \cdots, b_{\mathbf{v}-1}]$ be, respectively, the subband bit-allocation vector and the bit budget vector for all quality resolutions. The min-max approach for the problem (P) can be written as

$$\begin{aligned} &\min_{b,\beta} \max_{\mu} \mathcal{D}(\beta_0, \beta_1, \cdots, \beta_{\mathbf{v}-1}) \\ &\text{subject to } \sum_{z \in W} \beta_i[z] = b_i, \text{ for } i = 0, \cdots, \mathbf{v}-1, \\ &\text{and } \beta_{i-1}[z] \leq \beta_i[z], \ b_{i-1} \leq b_i \text{ for } i = 1, \cdots, \mathbf{v}-1. \end{aligned} \tag{38}$$

In other words, the min-max approach finds the best bit-allocation vectors for the preference distribution that yields the largest distortion.

First, we show that the least favorable preference distribution μ^* is independent of the quality resolution r. For any subband bit-allocation β_r at the quality resolution r, the least favorable preference distribution maximizes the distortion $\mathcal{D}(\beta_0, \beta_1, \cdots, \beta_{r-1}, \beta_r, \beta_r, \cdots, \beta_r)$. From Equation (29), we have

$$\mathcal{D}(\beta_0, ..., \beta_{r-1}, \beta_r, ..., \beta_r) \leq \max_{s=0,...,\mathbf{p}-1} \max_{t=0,\cdots,\mathbf{q}-1} \max_{k=0,\cdots,r} \Theta(s,t,\beta_k), \tag{39}$$

$$= \Theta(\mathbf{p}-1, \mathbf{q}-1, \beta_0), \tag{40}$$

where Equation (39) is derived from $0 \leq \mu_{(s,t,r)} \leq 1$, and $\sum_{(s,t,r)} \mu_{(s,t,r)} = 1$; and Equation (40) is obtained because the maximum distortion can be obtained when the bits β_0 for the coarsest quality resolution are assigned to the subbands in the highest spatial and temporal resolutions. Thus, the least favorable preference μ^* occurs when all users have preference 1 for the resolution $(\mathbf{p}-1, \mathbf{q}-1, 0)$.

After deriving the least favorable preference μ^*, the problem can be solved easily by using the methods proposed previously. It is noteworthy that the above min-max problem finds the optimal bit-allocation for the codec containing only one spatial, temporal, and quality resolution. This result corresponds to allocating the bits optimally for a non-scalable wavelet codec.

5. Experiment results

We now evaluate the coding performance of the proposed bit-allocation methods on a 2D+t wavelet encoder. In the experiment, a GOP has 32 frames. First, each frame is decomposed by applying a three-level 2-D wavelet transform with the 9-7 wavelets; then, the five-level MCTF method is applied to each spatial subband. The metod uses the 5-3 wavelets for temporal

decomposition of each spatial subband, as proposed in [17]. When MCTF is applied, we assume that the motion vectors are given. The motion estimation step uses a full search with integer-pixel accuracy. The block size is 16×16, and the search range is $[16, -15]$ both vertically and horizontally. Finally, the 2-D EZBC method [1, 8] is used to encode the wavelet coefficients of the 2D+t wavelet codec.

In the first two methods, the encoder knows each user's preference profile; and the third assumes that the preference profile is not available to the encoder. The first method is based on the Lagrangian approach, the second is based on the DP approach, and the third is based on the min-max approach.

We conduct four experiments on two video sequences: Foreman and Coastguard. Two of the experiments assume that there is only one quality resolution. The first experiment assumes that all the users subscribe to the same temporal resolution, but their spatial resolution preferences are different. Figure 6 shows the R-D curves of each sequence versus different spatial preference profiles. It is obvious that the scalable bit-allocation methods with known preferences achieve a better coding performance than the method that lacks the preference information. The DP method outperforms the Lagrangian method in all cases. Note that the DP method finds the optimal bit-allocation for the problem (P), while the Lagrangian method finds the optimal solution for the upper bound of the problem. In the second experiment, it is assumed that the only difference between the subscribers' preferences is the temporal resolution. That is, all users subscribe to the same spatial resolution. The experimental results shown in Figure 7 demonstrate that the DP method outperforms the Lagrangian method, and the min-max method has the worst performance in all cases. Note that the average $PSNR$ improvement of the DP method over the Lagrangian method in Figure 6 is higher than that in Figure 7, This indicates that knowing the preferences yields more $PSNR$ gain in the spatial resolution than in the temporal resolution.

The third experiment assumes that the users have three different spatial, temporal, and quality resolution preferences. In the experiment, there are three preference distribution settings. The preferences for the spatial, temporal, and quality resolutions are as follows: 1) temporal resolutions: 7.5 fps, 15 fps, and 30 fps; 2) spatial resolutions: $QuadCIF$, $QCIF$, and CIF; and 3) quality resolutions with bit constraints $b_0 = 2400$ kbps, $b_1 = 4600$ kbps, and $b_2 = 6200$ kbps.

Setting 1

$\mu_{r=2400kbit/GOP}$	**7.5**	**15**	**30 (fps)**
QuadQCIF	0.24	0	0
QCIF	0.06	0	0
CIF	0	0	0

$\mu_{r=4600kbit/GOP}$	**7.5**	**15**	**30 (fps)**
QuadQCIF	0	0.06	0
QCIF	0	0.24	0
CIF	0	0	0

$\mu_{r=6200kbit/GOP}$	**7.5**	**15**	**30 (fps)**
QuadQCIF	0	0	0
QCIF	0	0	0.08
CIF	0	0	0.32

http://dx.doi.org/10.5772/53427

Setting 2

$\mu_{r=2400kbit/GOP}$	**7.5**	**15**	**30 (fps)**
QuadQCIF	0.24	0.06	0
QCIF	0	0	0
CIF	0	0	0

$\mu_{r=4600kbit/GOP}$	**7.5**	**15**	**30 (fps)**
QuadQCIF	0	0	0
QCIF	0.06	0.24	0
CIF	0	0	0

$\mu_{r=6200kbit/GOP}$	**7.5**	**15**	**30 (fps)**
QuadQCIF	0	0	0
QCIF	0	0	0
CIF	0	0.08	0.32

Setting 3

$\mu_{r=2400kbit/GOP}$	**7.5**	**15**	**30 (fps)**
QuadQCIF	0.3	0	0
QCIF	0	0	0
CIF	0	0	0

$\mu_{r=4600kbit/GOP}$	**7.5**	**15**	**30 (fps)**
QuadQCIF	0	0	0
QCIF	0	0.3	0
CIF	0	0	0

$\mu_{r=6200kbit/GOP}$	**7.5**	**15**	**30 (fps)**
QuadQCIF	0	0	0
QCIF	0	0	0
CIF	0	0	0.4

The results are shown in Figure 8. Compared to Figures 6 and 7, there are no significant performance differences in the curves associated with the Lagrangian method and the min-max method.

6. Conclusion

We introduce the concept and the details of wavelet-based scalable video coding in this chapter. We also considers subscribers' preferred resolutions when assessing the performance of a wavelet-based scalable video codec. We formulate the problem as a scalable bit-allocation problem and propose different methods to solve it. Specifically, we show that the Lagrangian-based method can find the optimal solution of the upper bound of the problem, and that the dynamic programming method can find the optimal solution of the problem. We also consider applications where the subscribers' preferences are not known, and use a min-max approach to find a solution. Our experimental results show that knowing the users' preferences can improve the PSNR of the 2D+t wavelet scalable video codec. The average PSNR gain depends on the users' preference distribution. It can range from 1 db to 8 db at a fixed bit rate. There is a significant performance gap between when the preferences are known over when they are unknown. Hence, in our future work, we will reduce the gap or derive a method that enables use to estimate the preference patterns in the encoder.

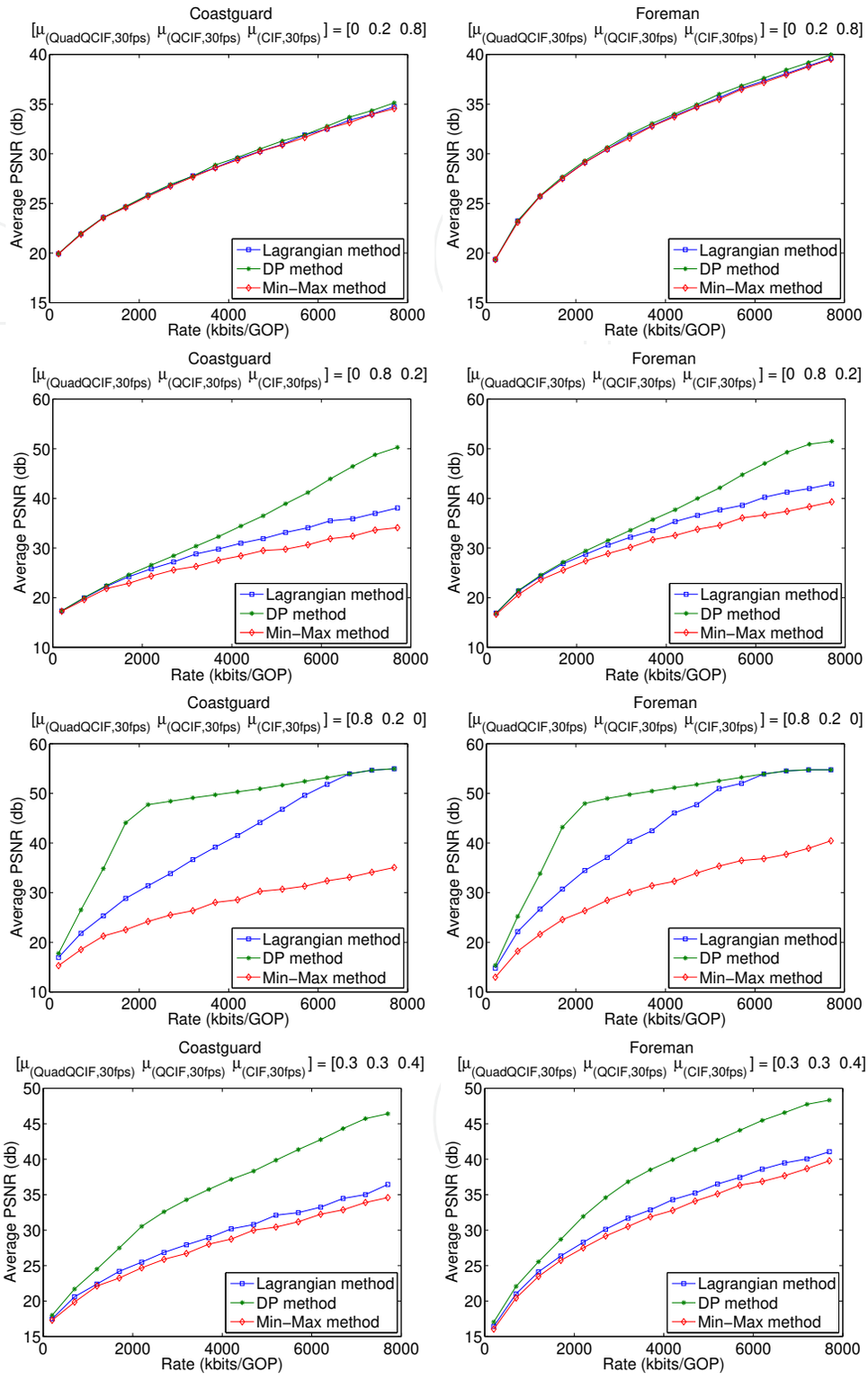

Figure 6. Comparison of the performance under different spatial preferences. Four preference patterns are used to subscribe to three spatial resolutions: QuadQCIF, QCIF, and CIF. Left: Coastguard. Right: Foreman. There is only one quality resolution and all users subscribe to 30 fps.

http://dx.doi.org/10.5772/53427

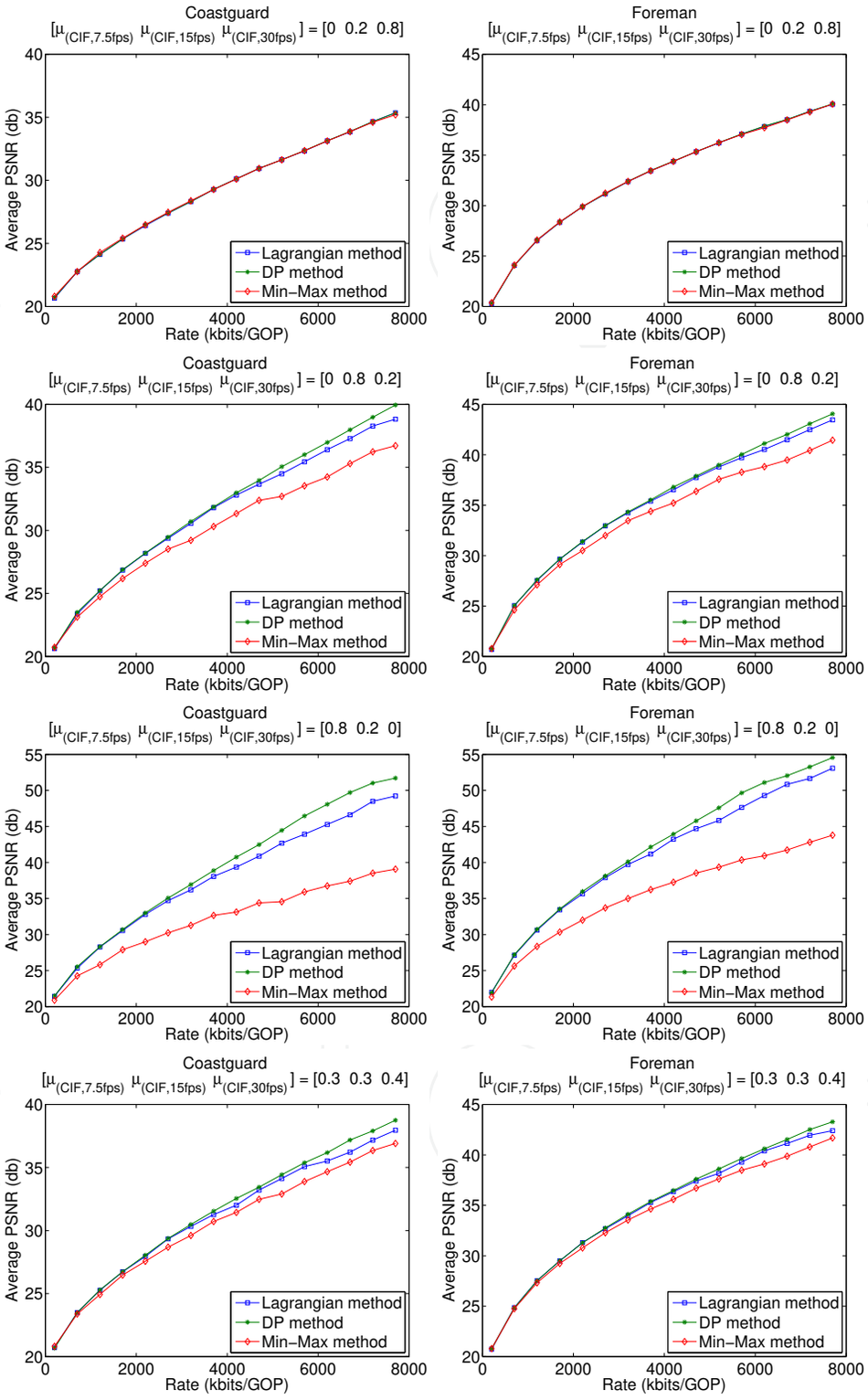

Figure 7. Comparison of the performance under different temporal preferences. Different preferences are used to subscribe to three temporal resolutions: 7.5 fps, 15 fps, and 30 fps. Left: Coastguard. Right: Foreman. There is only one quality resolution and all users subscribe to CIF.

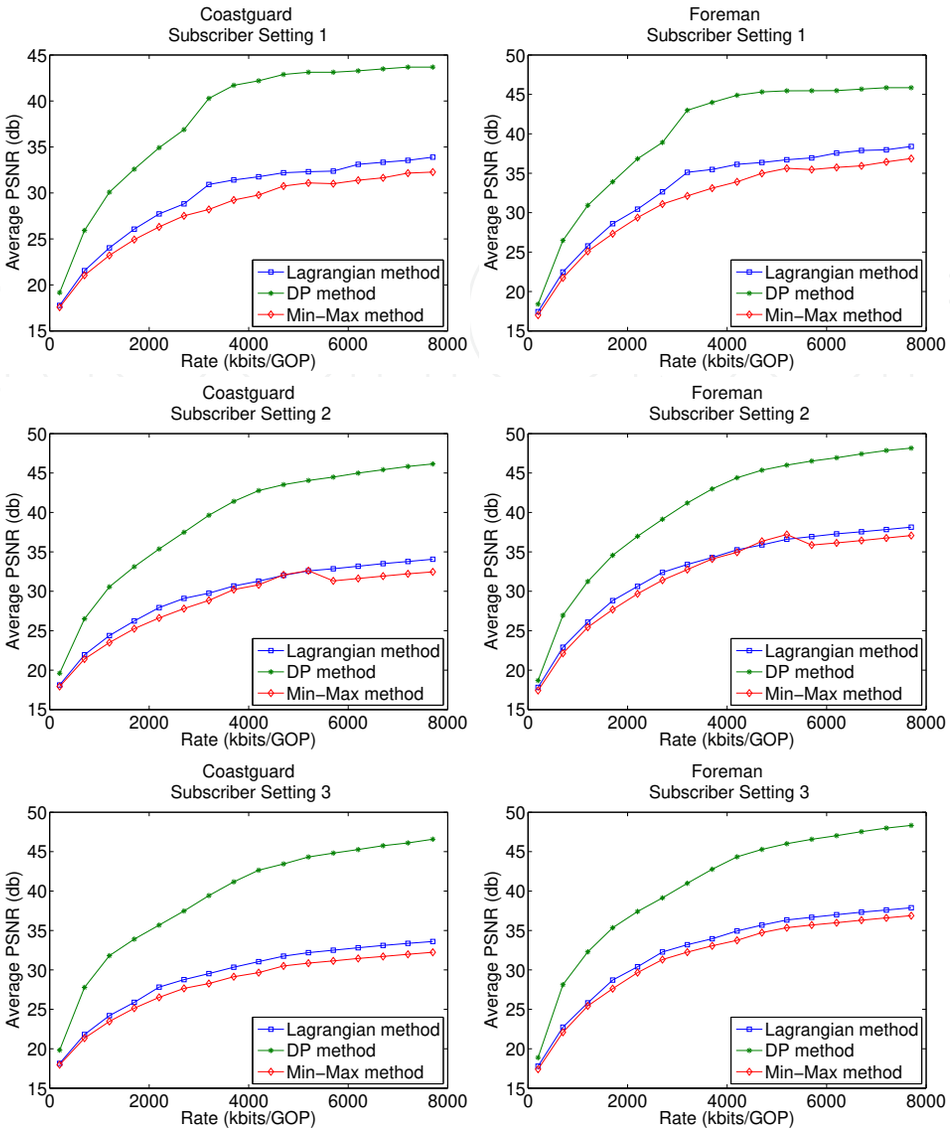

Figure 8. Comparison of the performance under different spatial, temporal, and quality preferences. Different preferences are used to subscribe to three temporal resolutions (7.5 fps, 15 fps, and 30fps), three spatial resolutions (QuadQCIF, QCIF, and CIF), and three quality resolutions (2400 kbps, 4600 kbps, and 6200 kbps). Left: Coastguard. Right: Foreman. Top row: the performance of setting 1; middle row: the performance of setting 2; and, bottom row: the performance of setting 3.

Author details

Guan-Ju Peng and Wen-Liang Hwang

Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan

References

[1] Shih-Ta Hsiang and J. W. Woods. Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank. *Signal Processing: Image Communications*, 16:705–724, May 2001.

[2] J. Barbarien M. Van der Schaar J. Cornelis Y. Andreopoulos, A. Munteanu and P. Schelkens. In-band motion compensated temporal filtering. *Signal Processing: Image Communications*, 19:653–673, August 2004.

[3] Qian Zhang, Q. Guo, Qiang Ni, Wenwu Zhu, and Ya-Qin Zhang. Sender-adaptive and receiver driven layered multicast for scalable video over the internet. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(4):482–495, April 2005.

[4] Heiko Schwarz, Detlev Marpe, and Thomas Wiegrand. Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1103–1120, 2007.

[5] Jens-Rainer Ohm. Three-dimensional subband coding with motion compensation. *IEEE Transactions on Image Processing*, 3(5):559–571, September 1994.

[6] Seung-Jong Choi and J. W. Woods. Motion-compensated 3-D subband coding of video. *IEEE Transactions on Image Processing*, 8(2):155–167, February 1999.

[7] K. Hanke T. Rusert and J.-R. Ohm. Transition filtering and optimized quantization in interframe wavelet video coding. *Proc. SPIE Visual Communications Image Processing*, 5150:682–694, 2003.

[8] Peisong Chen and John W. Woods. Bidirectional mc-ezbc with lifting implementation. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(10):1183–1194, October 2004.

[9] Jens-Rainer Ohm, Mihaela van der Schaar, and John W. Woods. Interframe wavelet coding - motion picture representation for universal scalability. *Signal Processing : Image Communication*, 19(9):877–908, 2004.

[10] Daniel Lee Michael J. Gormish and Michael W. Marcellin. JPEG 2000: Overview, architecture, and applications. In *IEEE International Conference on Image Processing*, pages 29–32, Sepetember 2000.

[11] Kannan Ramchandran, Antonio Ortega, and Martin Vetterli. Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders. *IEEE Transactions on Image Processing*, 3(5):533–545, 1994.

[12] Heiko Schwarz and Thomas Wiegrand. R-D optimized multi-layer encoder control for SVC. In *IEEE International Conference on Image Processing*, pages 281–284, Sepetember 2007.

[13] B. Usevitch. Optimal bit allocation for biorthogonal wavelet coding. In *Data Compression Conference*, pages 387 –395, March/April 1996.

[14] Cho-Chun Cheng, Guan-Ju Peng, and Wen-Liang Hwang. Subband weighting with pixel connectivity for 3-D wavelet coding. *IEEE Transactions on Image Processing*, 18(1):52–62, January 2009.

[15] Wen-Liang Hwang Guan-Ju Peng and Sao-Jie Chen. Fast implementation of the subband weighting for 3d wavelet coding. *ISRN Signal Processing*, 2011:Article ID 252734, 2011.

[16] David S. Taubman. High performance scalable image compression with EBCOT. *IEEE Transactions on Image Processing*, 9(7):1158–1170, 2000.

[17] Lin Luo, Jin Li, Shipeng Li, Zhenquan Zhuang, and Ya-Qin Zhang. Motion compensated lifting wavelet and its application in video coding. In *IEEE International Conference on Multimedia and Expo*, pages 365 – 368, August 2001.