

Article

Content Adaptive Lagrange Multiplier Selection for Rate-Distortion Optimization in 3-D Wavelet-Based Scalable Video Coding

Ying Chen and Guizhong Liu *

School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; xachenying@stu.xjtu.edu.cn

* Correspondence: liugz@xjtu.edu.cn; Tel.: +86-29-8266-5700

Received: 31 January 2018; Accepted: 6 March 2018; Published: 8 March 2018

Abstract: Rate-distortion optimization (RDO) plays an essential role in substantially enhancing the coding efficiency. Currently, rate-distortion optimized mode decision is widely used in scalable video coding (SVC). Among all the possible coding modes, it aims to select the one which has the best trade-off between bitrate and compression distortion. Specifically, this tradeoff is tuned through the choice of the Lagrange multiplier. Despite the prevalence of conventional method for Lagrange multiplier selection in hybrid video coding, the underlying formulation is not applicable to 3-D wavelet-based SVC where the explicit values of the quantization step are not available, with on consideration of the content features of input signal. In this paper, an efficient content adaptive Lagrange multiplier selection algorithm is proposed in the context of RDO for 3-D wavelet-based SVC targeting quality scalability. Our contributions are two-fold. First, we introduce a novel weighting method, which takes account of the mutual information, gradient per pixel, and texture homogeneity to measure the temporal subband characteristics after applying the motion-compensated temporal filtering (MCTF) technique. Second, based on the proposed subband weighting factor model, we derive the optimal Lagrange multiplier. Experimental results demonstrate that the proposed algorithm enables more satisfactory video quality with negligible additional computational complexity.

Keywords: rate-distortion optimization; mode decision; scalable video coding; Lagrange multiplier; 3-D wavelet-based SVC; motion-compensated temporal filtering

1. Introduction

With the rapid development of video services in recent years, how to efficiently compress video sequences has been considered as a very challenging task for transmitting video data over heterogeneous networks. To meet this demand, scalable video coding (SVC) has become a simple and flexible solution to enable seamless delivery by offering three different kinds of scalabilities, namely, temporal scalability, spatial scalability, and quality scalability [1–4]. Generally speaking, there are two main categories on SVC: discrete cosine transform (DCT)-based hybrid video coding SVC and wavelet-based SVC.

Owing to the intrinsic localization and multiresolution features of the discrete wavelet transform (DWT), video codecs based on motion-compensated three-dimensional (3-D) DWT have been studied extensively for use in SVC [5–8]. 3-D wavelet-based SVC provides a natural way in producing embedded bitstreams with full scalability and fine granularity for in-network adaptation [9,10]. In the 3-D wavelet-based SVC, temporal redundancy across frames is exploited by adopting the motion-compensated temporal filtering (MCTF) framework [11–14], and spatial redundancy inside a frame is utilized by 2-D spatial transform. Such codecs do not suffer from the drift problem often

exhibited by the DCT-based SVC codecs with their incorporated feedback loops. Consequently, over the last few years, much research has been devoted to the 3-D wavelet video coding, and it is meaningful to further explore its properties for developing a best coding approach for 3-D wavelet-based SVC.

As in hybrid video coding-based SVC, rate-distortion optimization (RDO) plays a significant role in SVC in choosing the most suitable coding parameters under communication bandwidth constraints. A critical problem residing in RDO is the selection of the Lagrange multiplier that controls the RD trade-off [15–23]. Sullivan and Wiegand first derived the relationship among the Lagrange multiplier, distortion and bitrate. This simple and effective formula has been extensively adopted in hybrid video coding [24,25]. Although the conventional Lagrange multiplier selection techniques have been developed and widely adopted in hybrid video coding, the existing formulations are not applicable to 3-D wavelet-based SVC. This is principally due to the fact that the Lagrange multiplier is determined by the quantization step. Different from hybrid video coding-based SVC where the quantization steps are implied in the base layer, quantization for wavelet-based SVC is usually approached in an embedded manner rather than using explicit quantization steps. Accordingly, in absence of the quantization step, the conventional Lagrange multiplier cannot be directly applied to 3-D wavelet-based SVC.

The Lagrange multiplier selection technique used to solve the RDO problem is an important feature that contributes to the success of current video coding. However, solving the optimal Lagrange optimization problem in 3-D wavelet coding is more complicated because the energy difference between the pixel and wavelet domains is not conserved after applying the bi-orthogonal wavelet transform and MCTF technique. The latter leads to the intrinsic subband coupling across various temporal subbands during motion aligned temporal filtering. Moreover, the open-loop prediction structure employed in 3-D wavelet-based SVC further complicates the issue of Lagrange multiplier selection. This issue has been solved by assigning different weighting factors to the subbands, resulting in degrading unpleasant quality fluctuations [18,19,26–28]. For the Haar filter used in temporal filtering structure, Ohm proposed a method to derive various quantization weights associated with low-frequency and high-frequency subbands [29]. Xiong et al. have obtained weighting factors for other longer temporal filters, such as 5/3 and 9/7 filters [30,31]. Peng et al. derived the weighting factors from the subscriber's preference for different resolutions [32,33]. In theory, the subband distortions presented in the reconstructed frame through inverse wavelet transform can be accurately derived from the filter-based weighting factors. However, all the temporal subbands in each temporal decomposition level have diverse features from each other. As a consequence, it is necessary to consider both the diverse content characteristics of the temporal subbands and the impact of subband coupling in the Lagrange multiplier selection process.

Inspired by the methods mentioned above, we propose a content adaptive Lagrange multiplier selection algorithm for RDO in 3-D wavelet-based SVC. During the RDO process, the wavelet filter types, subband coupling phenomenon, and temporal subband content information are all taken into account to adaptively compute the Lagrange multiplier. Our strategy aims to not only accurately select the Lagrange multiplier of each temporal subband in the MCTF decomposition level, but also yield better video quality. By performing the proposed algorithm, our codec clearly outperforms the existing well-known 3-D wavelet-based SVC coding schemes with higher PSNR gains and much lower video quality variations through the whole reconstructed video sequence.

The rest of this paper is organized as follows: Section 2 presents a brief overview of related work. The proposed content-adaptive Lagrange multiplier selection is introduced in detail in Section 3 with its performance evaluation provided in Section 4. Conclusions and discussions are drawn in Section 5.

2. Related Work

In this section, we review relevant background information related to the Lagrange multiplier selection in the wavelet-based SVC. The reasons for the failure of the conventional Lagrange multiplier selection in 3-D wavelet-based SVC are also addressed.

2.1. System Model

The representative motion compensated embedded zero block coding (MC-EZBC) scalable video coder [34] is a highly efficient member of the existing 3-D wavelet-based SVC schemes. However, the coding efficiency of MC-EZBC in the range of low bitrates is far from satisfactory. To improve this deficiency, Wu et al. put forward the well-known enhanced motion-compensated embedded zero block coding (ENH-MC-EZBC), which retains its excellent rate-distortion performance at high bitrates and achieves significant improvement at low bitrates and/or low resolutions [35].

In our work, the proposed algorithm is designed for the popular ENH-MC-EZBC considering quality scalability. Our choice of this codec system is motivated by the fact that ENH-MC-EZBC incorporates all the advanced encoding tools found in the state-of-the-art video coding schemes and obtains the excellent coding performance both at low and high bitrates. The ENH-MC-EZBC codec system model contains three parts: encoder, bitstream extractor, and decoder, which is shown in Figure 1. In the encoder, a motion compensated 3-D subband/wavelet transform naturally partitions the input video sequence into a range of spatiotemporal resolutions. We then use 3D-EZBC to encode the resulting spatio-temporal subbands. All the motion fields are coded as side information by using lossless spatial differential pulse code modulation (DPCM) and adaptive arithmetic coding. In the bitstream extractor, the fully embedded bitstream is truncated based on both user preferences and network conditions to generate a highly flexible scalable bitstream to meet specific applications. At the decoder, the respective reverse operations are carried out to reconstruct video sequences.

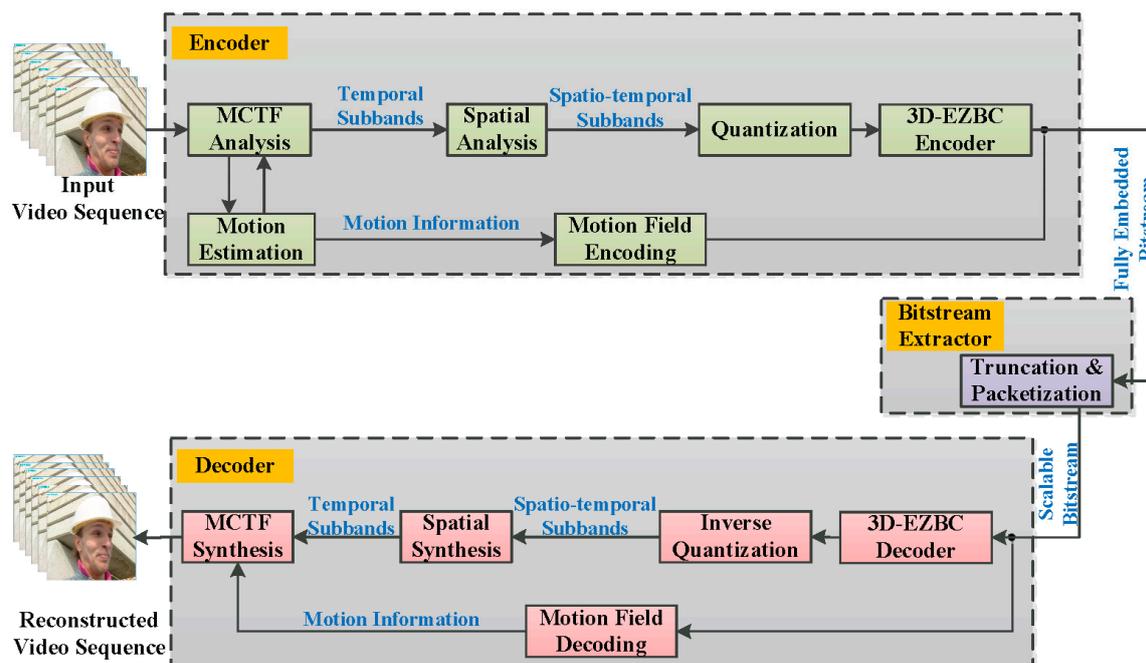


Figure 1. Block diagram of the ENH-MC-EZBC codec system model.

2.2. Lagrange Optimization and Lagrange Multiplier Selection

The Lagrange optimization technique provides a systematic way to solve the constrained RDO problem, which aims at selecting the optimal coding parameter that minimizes the overall distortion measure subject to a given target bitrate restriction.

More details on Lagrange optimization technique have been discussed in [16,20,36]. This technique is well known in optimization problems where the cost and objective functions are continuous and differentiable. Everett's contribution [37] demonstrated that the Lagrange optimization technique could also be used for discrete optimization problems, with no loss of optimality if a solution

exists with the required budget; i.e., as long as there exists a point in the convex hull that meets the required budget.

Let $\{\text{Para}\}$ denote the coding parameter set, including motion estimation, mode decision, MCTF decomposition levels, etc. For mathematical convenience, the rate control problem can be formulated as: Minimize the distortion D , subject to a target rate R_T :

$$\{\text{Para}\}_{\text{opt}} = \arg \min_{\{\text{Para}\}} D = \arg \min_{\{\text{Para}\}} \sum_{i=1}^K d_i \quad \text{s.t.} \quad \sum_{i=1}^K r_i \leq R_T, \quad (1)$$

where d_i and r_i are the distortion and bitrate for the i th ($i = 1, 2, \dots, K$) coding unit, respectively; K is the total number of coding units involved and R_T the available bitrate constraint.

In view of the Lagrange optimization, the above constrained optimization problem (1) can be converted into an unconstrained form as:

$$\{\text{Para}\}_{\text{opt}} = \arg \min_{\{\text{Para}\}} J = \arg \min_{\{\text{Para}\}} (D + \lambda R) = \arg \min_{\{\text{Para}\}} \sum_{i=1}^K (d_i + \lambda r_i), \quad (2)$$

where $J = D + \lambda R$ is the Lagrange cost function, and λ is the so-called Lagrange multiplier that weights the relative importance between d_i and r_i . The optimal coding parameter set for all coding units can be determined by minimizing the Lagrange cost function as expressed in Equation (2). Consequently, how to determine λ becomes a key problem in Lagrange optimization. To have a better solution to the unconstrained problem, much effort has been placed on the research of the Lagrange multiplier selection. One may attain the λ using bisection search [38,39]. In RDO for video coding, however, a more computationally efficient approach is usually favorable to determine the Lagrange multiplier.

Rather than empirically solving the problem of Lagrange multiplier selection as in [38,39], the λ in video coding can be determined using an R-D function. Due to the convexity of the R-D curve, the optimal slope λ matched to the desired R can be easily obtained using standard convex search techniques [24,40]. For each coding unit, the point on the R-D curve that minimizes the Lagrange cost is that point at which the line of absolute slope λ is tangent to the convex hull of the R-D curve.

Note that the R-D curve is convex and non-increasing in video coding, and if we assume that both R and D are differentiable everywhere, the minimum of the Lagrange cost function J is given by setting its derivative to zero, i.e.:

$$\frac{\partial J}{\partial R} = \frac{\partial D}{\partial R} + \lambda = 0, \quad (3)$$

which yields:

$$\lambda = - \left. \frac{\partial D}{\partial R} \right|_{R=R_T}. \quad (4)$$

where R_T is the target bitrate. A given value of λ yields an optimal solution $\{\text{Para}\}_{\text{opt}}(\lambda)$ to the original RDO problem (1) for a particular value of $R(\{\text{Para}\}_{\text{opt}})$.

2.3. Problem Formulation

In 3-D wavelet-based SVC, the temporal decomposition is efficiently refined by adopting the MCTF technique. The ENH-MC-EZBC scalable video coder utilizes an adaptive lifting-based MCTF framework with switching between LeGall and Tabatabai (LGT) 5/3 and Haar 2/2 filters to exploit the temporal redundancies between successive frames [41]. During the MCTF process, video frames are filtered into low-frequency (L) and high-frequency (H) subbands. The process to generate the temporal high- and low- subbands is called the “prediction” and “update” step, respectively. Figure 2 illustrates the prediction and updates steps for the analysis stage of the three-level adaptive MCTF decomposition. MC and IMC are motion compensation and inverse operators, respectively. Scene change information is

reflected in the choice of the filter bank, which propagates to the lower temporal levels. The coefficients α_τ and $\beta_\tau (\tau = 1, 2, 3)$ along the branches are filter-based weighting factors. Since 3-D wavelet-based SVC is usually based on the underlying open-loop MCTF structure, its RD performance is further complicated by the inherent problem of the propagation of the quantization errors along the temporal wavelet decomposition tree [42]. Therefore, the selection of the Lagrange multiplier for wavelet-based SVC should be addressed differently.

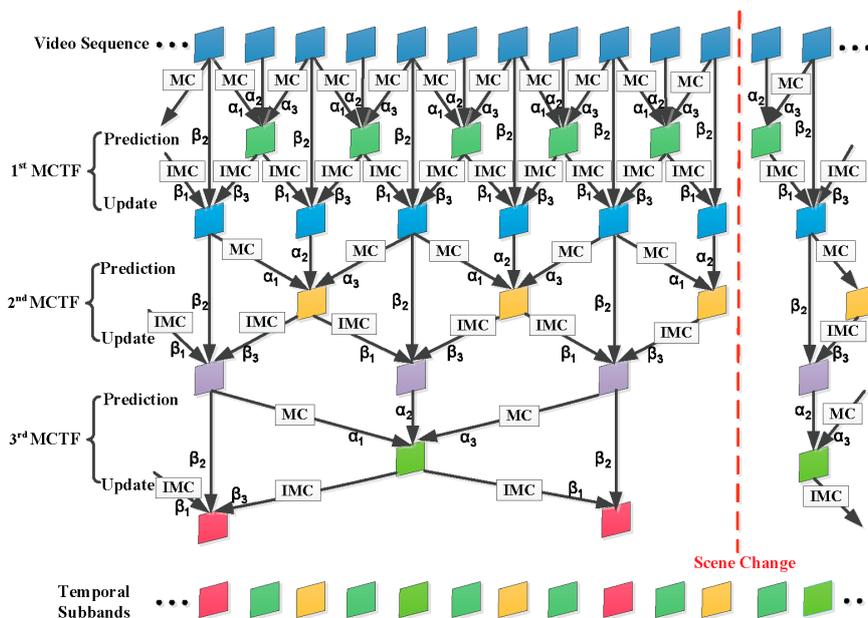


Figure 2. Lifting-based MCTF framework with adaptive switching based on Haar and 5/3 filters.

3. Proposed Lagrange Multiplier Selection Algorithm

In this section, we introduce the proposed content adaptive Lagrange multiplier selection algorithm for the 3-D wavelet-based SVC codec.

3.1. Lagrange Multiplier Selection Bottlenecks in 3-D Wavelet-Based SVC

Lagrange multiplier based mode decision is one of the most important technologies in SVC [43]. In the adaptive MCTF framework of 3-D wavelet-based SVC, each temporal subband frame in Figure 2 is assigned to one out of four frame modes, including “bi-direction” (denoted as m_1), “uni-left” (denoted as m_2), “uni-right” (denoted as m_3) and “intra” (denoted as m_4) modes. The ENH-MC-EZBC SVC codec incorporates an optional rate-distortion (R-D) optimized mode decision algorithm. The goal of an R-D optimized mode-selection algorithm is to choose the best mode from available coding frame modes. The mode that minimizes distortion subject to a rate constraint is chosen as the best frame mode. Let $F = (f_1, f_2, f_3, \dots, f_K)$ denote a group of K frames. For a vector of coding mode allocations $M = (m_1, m_2, m_3, m_4)$ and a bitrate constraint R_T , this optimization problem can be expressed by:

$$M^* = \arg \min_M D(F, M), \quad \text{s.t. } R(F, M) \leq R_T, \tag{5}$$

where M^* , $D(F, M)$ and $R(F, M)$ denote the vector of best mode allocations, distortion metric, and coded bitrate. This may be written as an unconstrained problem using a Lagrange optimization method as:

$$M^* = \sum_{i=1}^K \arg \min_M J(F_i), \tag{6}$$

where $J(F_i)$ is the Lagrange cost for the i th frame and is given as:

$$J(F_i) = D(F_i, M_i) + \lambda \cdot R(F_i, M_i), \tag{7}$$

where $D(F_i, M_i)$ is the distortion term between the current frame and its reference frame, and $R(F_i, M_i)$ the bitrate term representing the expected number of bits allocated to the i th frame, with a specific λ .

Accurate R-D model is crucial in the determination of the Lagrange multiplier. It is worthwhile to mention that the R-D relationships can be quite different for various temporal subbands (T-bands). Unfortunately, it cannot to maximize coding efficiency through utilizing a fixed Lagrange multiplier at each temporal layer. Consequently, the conventional Lagrange multiplier selection method is not optimal as it does not consider the content characteristics of the T-bands.

3.2. Analyzing the Distortion Relationship between Temporal Subbands and Reconstructed Frames

The distortion fluctuation exhibited by the 3-D wavelet-based SVC codecs can be better understood by analyzing the distortion propagation during MCTF. In the ENH-MC-EZBC codec, the adaptive MCTF framework has been implemented using either the Haar or the 5/3 filter to improve the coding performance. As shown in Figure 2, the coefficients α_τ and β_τ ($\tau = 1, 2, 3$) along the branches are filter-based weighting factors, which are given in Table 1.

Table 1. Coefficients α_τ and β_τ depending on the mode of the current frame.

Frame Mode	α_1	α_2	α_3	β_1	β_2	β_3
bi-direction	-1/2	1	-1/2	1/4	1	1/4
uni-left	-1	1	0	1/2	1	0
uni-right	0	1	-1	0	1	1/2
intra	N/A	N/A	N/A	N/A	1	N/A

In the Haar MCTF, the original frames are filtered temporally with a two-tap Haar filter along the motion trajectory. For the connected pixels, the low-frequency subbands and high-frequency subbands are implemented by the following lifting structure:

$$H_i(m, n) = I_{2i+1}(m, n) - \tilde{I}_{2i}(m - d_m, n - d_n), \tag{8}$$

$$L_i(m - \bar{d}_m, n - \bar{d}_n) = I_{2i}(m - \bar{d}_m, n - \bar{d}_n) + (1/2) \cdot \tilde{H}_i(m - \bar{d}_m + d_m, n - \bar{d}_n + d_n). \tag{9}$$

For the 5/3-based MCTF, the temporal analysis can be written as:

$$H_i(m, n) = I_{2i+1}(m, n) - (1/2) \cdot \tilde{I}_{2i}(m - d_m, n - d_n) - (1/2) \cdot \tilde{I}_{2i+2}(m + d_m, n + d_n), \tag{10}$$

$$L_i(m - \bar{d}_m, n - \bar{d}_n) = I_{2i}(m, n) + (1/4) \cdot \tilde{H}_i(m - \bar{d}_m - d_m, n - \bar{d}_n - d_n) + (1/4) \cdot \tilde{H}_{i-1}(m - \bar{d}_m + d_m, n - \bar{d}_n + d_n), \tag{11}$$

where H_i and L_i denote the temporal high- and low-frequency subbands of the i th ($i = 1, 2, \dots, K$) frame of a video sequence, respectively; $I_{2i}(m, n)$ represents the pixel value at position (m, n) in frame f_{2i} ; $\tilde{I}_{2i}(m - d_m, n - d_n)$ and $\tilde{I}_{2i+2}(m + d_m, n + d_n)$ denote the interpolated values of pixel (m, n) in the frame $2i$ and $2i + 2$, respectively. \tilde{H}_i is the interpolated value of pixel in the high-frequency frame; (d_m, d_n) is the motion vector in frame f_{2i} ; \bar{d}_m is the closest integer value to d_m , and \bar{d}_n is defined in the same way.

For mathematical convenience, we only consider the case in the lifting structure of one-level 5/3 inverse MCTF in the following. Thus, according to the Equations (10) and (11), the errors in the reconstructed $f_{2i}(m, n)$ and $f_{2i+1}(m, n)$ can be formulated as:

$$\varepsilon_{f_{2i}} = -(1/4) \cdot \varepsilon_{H_{i-1}} + 1 \cdot \varepsilon_{L_i} - (1/4) \varepsilon_{H_i}, \tag{12}$$

$$\varepsilon_{f_{2i+1}} = -(1/8) \cdot \varepsilon_{H_{i-1}} + (1/2) \cdot \varepsilon_{L_i} + (3/4) \cdot \varepsilon_{H_i} + (1/2) \cdot \varepsilon_{L_{i+1}} - (1/8) \cdot \varepsilon_{H_{i+1}}, \tag{13}$$

where ε_f denotes the reconstruction error of frame f . The assumption supposes that the distortion in each T-band can be modeled as independent and identically-distributed zero-mean additive white noise. Let σ_L^2 and σ_H^2 be the error variance of pixels in low-frequency subbands and that of pixels in high-frequency subbands, respectively. Also, let $\sigma_{f_{2i}}^2$ and $\sigma_{f_{2i+1}}^2$ denote the error variance in the reconstructed video frames. Accordingly, the distortions in the reconstructed frames are expressed as:

$$\sigma_{f_{2i}}^2 = 1 \cdot \sigma_L^2 + (1/8) \cdot \sigma_H^2, \quad \sigma_{f_{2i+1}}^2 = (1/2) \cdot \sigma_L^2 + (19/32) \cdot \sigma_H^2. \tag{14}$$

The distortion associated with different T-bands will contribute differently to distortion in the reconstructed frames. Typically, this difference is quantified by analyzing the weighting factor for inverse MCTF. The weighting factor indicates how much a unit distortion in a specified subband contributes to the overall distortion in the reconstructed video. The derivation of its weighting factors is given as:

$$\omega_L = (1/2)^2 + (1)^2 + (1/2)^2 = 1.5, \tag{15}$$

$$\omega_H = (-1/8)^2 + (-1/4)^2 + (3/4)^2 + (-1/4)^2 + (-1/8)^2 \approx 0.7188. \tag{16}$$

This results in distortion fluctuation across the reconstructed frames after one-level inverse MCTF. When multi-level MCTF is involved, the weighting factor of each temporal subband is modeled as:

$$\omega_{LL} = \omega_L \cdot \omega_L, \quad \omega_{LH} = \omega_L \cdot \omega_H, \quad \omega_{LLL} = \omega_{LL} \cdot \omega_L, \quad \omega_{LLH} = \omega_{LL} \cdot \omega_H, \dots \tag{17}$$

Therefore, the distortion relationship between the T-bands and the reconstructed frame can be derived in the same way but with an iterative calculation. That is, the distortion for the original video sequences (denoted as D) after T levels MCTF decomposition can be derived by:

$$D = \sum_{i=1}^K D_i = \sum_{i=1}^K \left(\sum_{t=1}^T \prod_{j=1}^{t-1} \omega_{i-H}^{(j)} d_{i-H}^{(t)} + \prod_{t=1}^T \omega_{i-L}^{(t)} d_{i-L}^{(T)} \right), \tag{18}$$

where $d_{i-H}^{(t)}$, and $d_{i-L}^{(T)}$ are the reconstruction distortion (namely error variance) of the high-frequency subbands in the t th ($1 \leq t \leq T$) temporal level, and low-frequency subbands in the highest temporal level, respectively. Additionally, $\omega_{i-H}^{(j)}$ and $\omega_{i-L}^{(t)}$ represent the weighting factors dependent on the motion estimation algorithm and the wavelet filter pair used for MCTF.

3.3. Adaptive Lagrange Multiplier Selection

As described in Subsection 2.2, the Lagrange multiplier λ is usually used to guide the bit allocation of each subband so that the overall distortion can be minimized and also with balanced the rate and distortion among the reconstructed frames. To our best knowledge, a larger λ will result in a higher coding distortion with less coding bits. Contrarily, a smaller λ will lead to a lower coding distortion with more coding bits. Therefore, we can allocate more bits to the T-bands with more detailed information by decreasing the subband-level Lagrange multiplier.

We define a weighting factor vector as $\Omega = (\omega_1, \omega_2, \dots, \omega_K)$, which indicates the contribution of a unit quantization error in each temporal subband to the overall distortion in the reconstructed video sequence. To obtain the weighting factor of the MCTF process, we need to employ a novel derivation, which involves the complicated motion compensation process and temporal subband content characteristics statistics. For computing these weighting factors, we utilize the mutual information, gradient per pixel, and texture homogeneity to measure the temporal subband content features.

Mutual information (MI) is a measurement of similarity between frames which can detect the differences among successive frames [44]. In this paper, we utilize MI to measure the similarity between the temporal subband frames. A large difference between frames (corresponding to the high motion activity) leads to a low MI value, while a small change between frames responds to a high MI value [45,46].

Let X be a discrete random variable with a set of possible outcomes $A_X = \{a_1, a_2, \dots, a_L\}$ with possibilities $\{p_1, p_2, \dots, p_L\}$, $p_X(x = a_l) = p_{ll} \geq 0$, and $\sum_{x \in A_X} p_X(x) = 1$. According to the information theory, the entropy of X is:

$$H(X) = - \sum_{x \in A_X} p_X(x) \log p_X(x). \quad (19)$$

The joint entropy of discrete random variables X and Y is:

$$H(X, Y) = - \sum_{x, y \in A_X, A_Y} p_{XY}(x, y) \log [p_{XY}(x, y)]. \quad (20)$$

The MI between random variable X and Y is given by:

$$I(X, Y) = - \sum_{x, y \in A_X, A_Y} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)}. \quad (21)$$

The relation between MI and joint entropy is given by:

$$I(X, Y) = H(X) + H(Y) - H(X, Y). \quad (22)$$

In a YUV formatted video sequence, let us consider a gray level video sequence with intensity value ranging from 0 to $L - 1$ (e.g., $L = 256$ for 8-bit depth). For the luminance component Y , $P_{i,i+1}^Y(x, y)$ ($0 \leq x, y \leq L - 1$) is the probability that a pixel with gray level x in frame f_i has a gray level y in frame f_{i+1} . Let $MI_{i,i+1}^Y$ be the MI of the luminance component. So we can obtain the $MI_{i,i+1}^Y$ value as shown below:

$$MI_{i,i+1}^Y = - \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} P_{i,i+1}^Y(x, y) \log \left[\frac{P_{i,i+1}^Y(x, y)}{P_i^Y(x)P_{i+1}^Y(y)} \right], \quad (23)$$

where $P_i^Y(x)$ and $P_{i+1}^Y(y)$ are the possibilities that a pixel with gray level x in frame f_i and a pixel with gray level y in frame f_{i+1} , respectively.

The gradient based subband content complexity measure has been considered in our Lagrange multiplier selection scheme. Here, the gradient per pixel (GPP) of the temporal subband is defined by:

$$GPP_i = \frac{1}{M \times N} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} (|I_i(m, n) - I_i(m+1, n)|) + (|I_i(m, n) - I_i(m, n+1)|), \quad (24)$$

where M and N are the width and height of the temporal subband, and $I_i(m, n)$ is the pixel value at position (m, n) .

With the statistical analysis, we observe that, if the subband belongs to the homogeneous regions, the number of bits will be a smaller value, and it will relevantly increase with the complex of the texture. Hence, the texture homogeneity of each subband is measured by calculating the ratio between variance and mean of the subband, which is given by:

$$\text{Ratio}_k = \frac{\text{Variance}}{\text{Mean}} = \frac{\text{Variance}}{\frac{1}{256} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} I_i(m, n)}, \tag{25}$$

where Variance and Mean are the variance and mean value of each temporal subband, respectively.

Let W denote the synthesis gain matrix with regard to target video sequence. Therefore, the associated temporal subband weighting factor $\Omega = (\omega_1, \omega_2, \dots, \omega_K)$ for prediction can be calculated as:

$$\omega_i = \omega_{\text{filter}_i} \cdot (\alpha \cdot \text{GPP}_i + \beta \cdot \text{Ratio}_i)^{-\gamma \cdot \text{MI}_i} \cdot W_{i \rightarrow i+1}, \tag{26}$$

where α , β , and γ are model parameters updated by multiple linear regression analysis. For each temporal subband frame i , ω_{filter_i} is the filter-based weighting factor using Equations (15)–(17); MI_i , GPP_i , and Ratio_i can be computed by Equations (23)–(25), respectively. As pointed out in [31], $W_{i \rightarrow i+1}$ is the cross-subband error propagation, which reflects the importance of coefficients in each subband.

Similar to JPEG2000 [47], we assume that the distortions from various temporal subbands are approximately additive. Equation (18) reveals that the total distortion is simply a linear combination of the distortions of all the temporal subbands. It is noticeable that the wavelet coefficients within each temporal subband obey the Gaussian distribution [48,49], hence, the distortion of the transform coefficients can be expressed as follows:

$$d_i = k' \frac{\pi e}{6} \left(\prod_{q=1}^Q \sigma_q^2 \right)^{1/Q} 2^{-\eta r_i} \omega_i, \tag{27}$$

where d_i and r_i stand for the distortion and bitrate for the i th temporal subband frame, respectively. σ_q^2 is the variance of the q th coefficient, Q the number of coefficients concerned. Besides, k' and η are the parameters related to the temporal decomposition scheme used and the corresponding slope for SVC, respectively.

However, in practice, it is very difficult to determine the variance for a given coefficient, as required by Equation (27). Based on the observation the variance of each coefficient can be estimated by:

$$\sigma_q^2 = \beta_q \sigma_f^2, \tag{28}$$

where β_q is a parameter related to the wavelet transform, and σ_f^2 the variance of the residual pixel values before wavelet transform. Therefore, Equation (27) can then be rewritten as:

$$d_i = k' \cdot \frac{\pi e}{6} \cdot \left(\prod_{q=1}^Q \beta_q \sigma_f^2 \right)^{1/Q} \cdot 2^{-\eta r_i} \cdot \omega_i. \tag{29}$$

Moreover, the variance of the residual pixel value before wavelet transform σ_f^2 can be further approximated using the mean absolute difference (MAD) by $\sigma_f^2 \approx 2\text{MAD}^2$. Consequently, the distortion d_i for the i th subband can be expressed as:

$$\begin{aligned} d_i &= k' \cdot \frac{\pi e}{6} \cdot \left(\prod_{q=1}^Q \beta_q \sigma_f^2 \right)^{1/Q} \cdot 2^{-\eta r_i} \cdot \omega_i \\ &= k' \cdot \frac{\pi e}{6} \cdot \left(\prod_{q=1}^Q \beta_q \cdot 2\text{MAD}^2 \right)^{1/Q} \cdot 2^{-\eta r_i} \cdot \omega_i \\ &= k \cdot \frac{\pi e}{6} \cdot \text{MAD}^2 \cdot 2^{-\eta r_i} \cdot \omega_i \end{aligned} \tag{30}$$

with $k = 2k' \left(\prod_{q=1}^Q \beta_q \right)^{1/Q}$ a parameter related to the wavelet transform decomposition scheme used.

According to Equations (26)–(30), it is easy to compute the Lagrange multiplier λ_i of the i th subband as:

$$\lambda_i = - \left. \frac{\partial d_i}{\partial r_i} \right|_{r_i=R_i}, \quad (i = 1, 2, \dots, K), \tag{31}$$

Correspondingly, Equation (31) can be written as:

$$\begin{cases} \lambda_1 = -\omega_1 \cdot 2^{-\eta} \cdot \left. \frac{\partial d_1}{\partial r_1} \right|_{r_1=R_1} \\ \lambda_2 = -\omega_2 \cdot 2^{-\eta} \cdot \left. \frac{\partial d_2}{\partial r_2} \right|_{r_2=R_2} \\ \vdots \\ \lambda_K = -\omega_K \cdot 2^{-\eta} \cdot \left. \frac{\partial d_K}{\partial r_K} \right|_{r_K=R_K} \end{cases}, \tag{32}$$

Based on Equations (30)–(32), a more practical Lagrange multiplier expression is then derived as:

$$\lambda_i = \varphi \text{MAD}^2 2^{-\eta r_i} \omega_i, \tag{33}$$

with $\varphi = 6.2$ an empirical constant suitable for different sequences.

By means of Lagrange multiplier λ_i , the RDO objective function Equation (7) can be rewritten as:

$$\begin{aligned} J(F_i) &= D_i(F_i, M_i) + \lambda_i \cdot R_i(F_i, M_i) \\ &= \sum_{i=1}^K \omega_i d_i(r_i) + \lambda_i \left(\sum_{i=1}^K R_i - R_T \right) \\ &= \sum_{i=1}^K \left[\sum_{t=1}^T \prod_{j=1}^{t-1} \omega_{i-H}^{(j)} d_{i-H}^{(t)}(r_i) + \prod_{t=1}^T \omega_{i-L}^{(t)} d_{i-L}^{(T)}(r_i) \right] + \lambda_i \cdot \left(\sum_{i=1}^K R_i - R_T \right) \end{aligned}, \tag{34}$$

where $\omega_{i-H}^{(j)} d_{i-H}^{(t)}(R_i)$ denote the weighted distortion term for the high-frequency subbands, and $\omega_{i-L}^{(t)} d_{i-L}^{(T)}(R_i)$ the weighted distortion term for the low-frequency subbands. Only the frame mode with the minimum Lagrange cost $J(F_i)$ is finally chosen as the best frame mode.

3.4. Summary of the Proposed Algorithm

The detailed procedure of the proposed Lagrange multiplier selection algorithm is outlined in the following Algorithm 1.

In order to access the accuracy of the proposed $R - \lambda$ model, R^2 is utilized as the quantitative metric which can measure the degree of data variation from a given model [50]:

$$R^2 = 1 - \frac{\sum_i (X_i - \hat{X}_i)^2}{\sum_i (X_i - \bar{X}_i)^2}, \tag{35}$$

where X_i and \hat{X}_i , respectively, are the actual and the estimated values of the i th data point. \bar{X} is the mean of all the data points. The maximum R^2 value is 1, which occurs when $X_i = \hat{X}_i$ for any i . The closer the value of R^2 is to 1, the more accurate the model is. In our experiment, for each test sequence, the λ value of each MCTF level is fitted at the target bitrate by the proposed $R - \lambda$ model. The R^2 statistics for four-level MCTF are tabulated in Table 2. From this table, we can notice that the R^2 values are all close to 1. Consequently, it can be concluded that the analytical $R - \lambda$ model works well for sequences at the different temporal decomposition levels. This is a crucial point of our approach.

Algorithm 1 An efficient algorithm for content adaptive Lagrange multiplier selection

Step 1 Initialization: Input a video sequence with K frames for illustration. Obtain the frame rate, total number of frames, target bitrate, etc., from the configuration file or manual input.

Step 2 Modeling the distortion relationship between the temporal subbands and reconstructed frames.

Step 3 Temporal subband weighting factor calculation (for temporal subband i)

(3-1) Calculate MI value $MI_{i,i+1}^Y$ of luminance component between f_i and f_{i+1} by Equation (23);

(3-2) GPP_k calculation for each subband according to Equation (24);

(3-3) Texture homogeneity $Ratio_k$ calculation for each subband according to Equation (25);

(3-4) Apply $MI_{i,i+1}^Y$, GPP_k , and $Ratio_k$ into Equation (26) to compute temporal subband weighting factor ω_i .

Step 4 Lagrange multiplier selection: Carry out the adaptive Lagrange multiplier selection scheme by means of Equations (30)–(33).

Step 5 Model parameters update: Parameters update using multiple linear regression analysis.

Step 6 Temporal subband frame mode selection:

(6-1) Calculate the Lagrange cost $J(F_i)$ for the temporal subband frame i with mode $M = (m_1, m_2, m_3, m_4)$.

(6-2) Output the best frame mode M^* with the minimum cost $J(F_i)$.

Step 7 Loop until all frames are encoded:

Set $i := i + 1$, if $i < K$, going to **Step 2** until the end of encoding, else report the related encoding parameters and exit.

Table 2. The R^2 values of each MCTF level for test video sequences.

Sequences	MCTF Level			
	1st	2nd	3rd	4th
Football	0.9921	0.9894	0.9951	0.9975
Foreman	0.9860	0.9954	0.9967	0.9982
Soccer	0.9984	0.9962	0.9981	0.9993
Crew	0.9976	0.9983	0.9989	0.9991
Ice	0.9787	0.9899	0.9935	0.9980
City	0.9843	0.9894	0.9932	0.9982
Johnny	0.9885	0.9891	0.9964	0.9989
KristenAndSara	0.9932	0.9967	0.9970	0.9985
Stockholm	0.9924	0.9962	0.9984	0.9990
Basketball	0.9787	0.9812	0.9899	0.9988
Cactus	0.9949	0.9950	0.9966	0.9990
Park_joy	0.9815	0.9843	0.9957	0.9981
Traffic	0.9870	0.9932	0.9960	0.9991
PeopleOnStreet	0.9893	0.9919	0.9949	0.9993
Average	0.9888	0.9919	0.9957	0.9986

4. Experimental Results

In this section, extensive experiments have been conducted to verify the effectiveness of the proposed content adaptive Lagrange multiplier selection algorithm for 3-D wavelet-based SVC.

4.1. Video Test Sequences

The experimental results over all video coding schemes are reported for fourteen standard test sequences based on YUV color format with 4:2:0 color sampling and 8 bits of precision per sample,

including five different resolutions: CIF (352 × 288, 30 fps), 4CIF (704 × 576, 30 fps), 720p (1280 × 720, 60 fps), 1080p (1920 × 1080, 50 fps), and 2K (2560 × 1600, 30 fps), which are listed in Table 3. We also plot the temporal information (TI) and spatial information (SI) indices [51,52] of all source sequences in Figure 3. It demonstrates that the test sequences cover a wide range of video contents in terms of motion and spatial details. For each of the resolutions, the test sequences are decoded at different target bitrates (CIF and 4CIF: 256, 384, 512, 640, 768, 896, 1024 kbps; 720p: 384, 512, 640, 768, 896, 1024, 1536 kbps; 1080p and 2K: 2048, 3072, 4096, 5120, 6144, 7168, 10240 kbps).

Table 3. Properties of standard video test sequences.

Sequences	Resolution	K ¹	Characteristics
Football	352 × 288	260	Fast camera and human subject motion, highly spatial details
Foreman	352 × 288	300	Fast camera and content motion with pan at the end
Soccer	352 × 288	300	Fast changes in motion, rapid camera panning
Crew	704 × 576	300	Multiple moderate objects movement
Ice	704 × 576	240	Still background and moderate human subject motion
City	704 × 576	300	Fast camera motion, high detail of buildings
Johnny	1280 × 720	100	Still background and low local motion
KristenAndSara	1280 × 720	100	Still background and moderate local motion
Stockholm	1280 × 720	100	Moderate camera panning, high detail of buildings
Basketball	1920 × 1080	100	Fast camera and human subject motion, highly spatial details
Cactus	1920 × 1080	100	Circling motion and highly spatial details
Park_joy	1920 × 1080	100	Camera and content motion, high detail of trees
Traffic	2560 × 1600	100	Moderate translational motion and highly spatial details
PeopleOnStreet	2560 × 1600	100	Still background and many human subject motion

¹ The number of frames in the test video sequence.

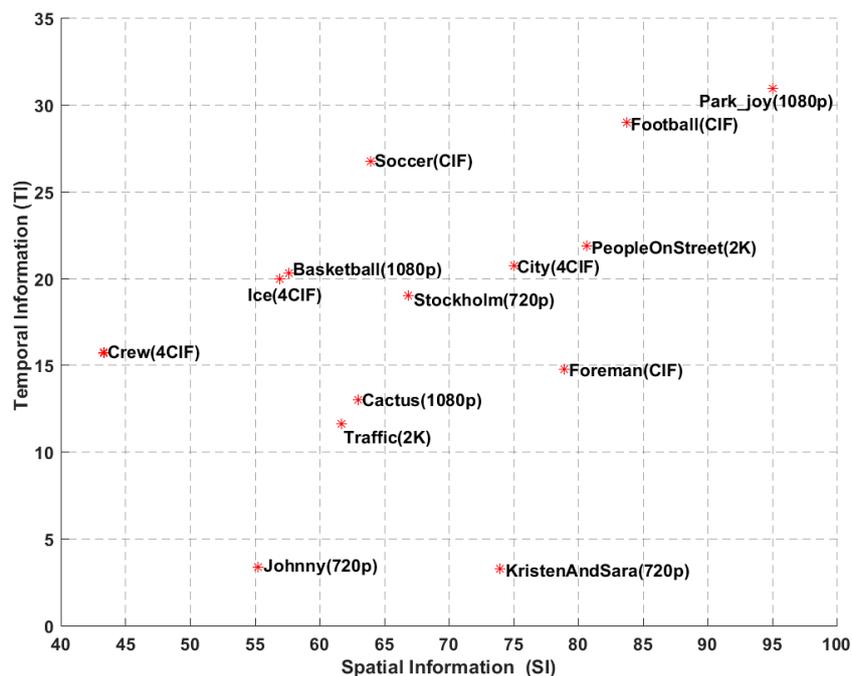


Figure 3. The spatial and temporal information indices of the test sequences (red star represents the coordinate value of SI and TI in the test sequence).

4.2. Experimental Setup

In this paper, all the algorithms are implemented with ANSI C in Microsoft Visual C++ 6.0 and MATLAB R2012b programming environments. Our experiments are conducted on a 4-core (i5-2400@3.10GHz) computer equipped with RAM 8 GB that is also used to measure the computational

complexity of our method. In the simulation, each group of pictures (GOP) contains 16 frames. During the MCTF process, motion estimation is implemented using a full-search with quarter-pixel accuracy on the dyadic wavelet coefficients. The block size varies from 4×4 to 128×128 , and the search range of both the horizontal and vertical dimension is $[-16, 15]$. The default Lagrange multiplier values are given in the rate-distortion optimized mode selection process. All the other encoder settings are set identically for all methods.

4.3. Performance Evaluation

To evaluate the effectiveness of our algorithm, we have implemented it on the 3-D wavelet-based SVC reference software ENH-MC-EZBC configured with the common test conditions as suggested in configure file. In the simulation, we have performed the following five representative codecs with the same configuration. Recently, these codecs deliver the best coding performance for scalable coding of video datasets [53]:

- MC-EZBC: the original MC-EZBC without employing proper RDO scheme.
- RPI-MC-EZBC: the bidirectional MC-EZBC from Rensselaer Polytechnic Institute which uses Haar filters for the conventional MCTF framework with the default Lagrange multiplier value for all the temporal decomposition levels.
- RWTH-MC-EZBC: the improved version of MC-EZBC from RWTH Aachen University, which uses longer filters instead of Haar filters for the conventional MCTF framework with the corresponding fixed Lagrange multiplier for each temporal decomposition level.
- ENH-MC-EZBC: the enhanced MC-EZBC using an adaptive MCTF framework with the corresponding fixed Lagrange multiplier for each temporal decomposition level.
- Proposed method: our codec with the proposed content adaptive Lagrange multiplier selection method.

In the experiments, only the luminance component is taken into consideration since human visual system is less sensitive to color than to luminance. For reasons of brevity, the average peak signal-to-noise ratios (PSNR) (dB) and the standard deviation of PSNR (PSNR STD) on luminance component have been used as quality metric, which are defined as follows:

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right), \quad (36)$$

$$\text{PSNR STD} = \sqrt{\frac{K \cdot \sum_{i=0}^{K-1} \text{PSNR}_i^2 - \left(\sum_{i=0}^{K-1} \text{PSNR}_i \right)^2}{K \cdot (K - 1)}}, \quad (37)$$

where MSE denotes the mean square error between the original frame and reconstructed frame, PSNR_i stands for the PSNR of the i th reconstructed frame and K the number of video frames in a sequence. Note that a higher PSNR means that a better RD performance is achieved. Meanwhile, the smaller the PSNR STD value, the better the video quality perceived by the end user, and vice versa.

4.3.1. Comparison of Rate-Distortion Performance

To verify the overall rate-distortion (R-D) performance of the proposed Lagrange multiplier selection approach, we compare it with the successful coding schemes on the framework of motion-compensated subband coding (MCSBC): ENH-MC-EZBC, RWTH-MC-EZBC, RPI-MC-EZBC, and MC-EZBC. Let "Scheme 1", "Scheme 2", "Scheme 3" and "Scheme 4" denote our method compared to the ENH-MC-EZBC, RWTH-MC-EZBC, RPI-MC-EZBC, and MC-EZBC, respectively.

Figure 4 shows the R-D curves of five codecs for six selected test sequences ("Soccer", "Crew", "Stockholm", "Basketball", "Park_joy", and "PeopleOnStreet") at different target bitrates. Obviously, it

is observed that the codec with the proposed algorithm achieves the best R-D performance among all codecs. From these figures, we can see that our method is noted with average PSNR improvement over other methods about 0.53–3.2 dB. As shown in Figure 4a, for the “Soccer” sequence, the proposed algorithm yields 0.79, 1.15, 1.4, and 2.7 dB higher PSNR values than the ENH-MC-EZBC, RWTH-MC-EZBC, RPI-MC-EZBC, and MC-EZBC in average, respectively.

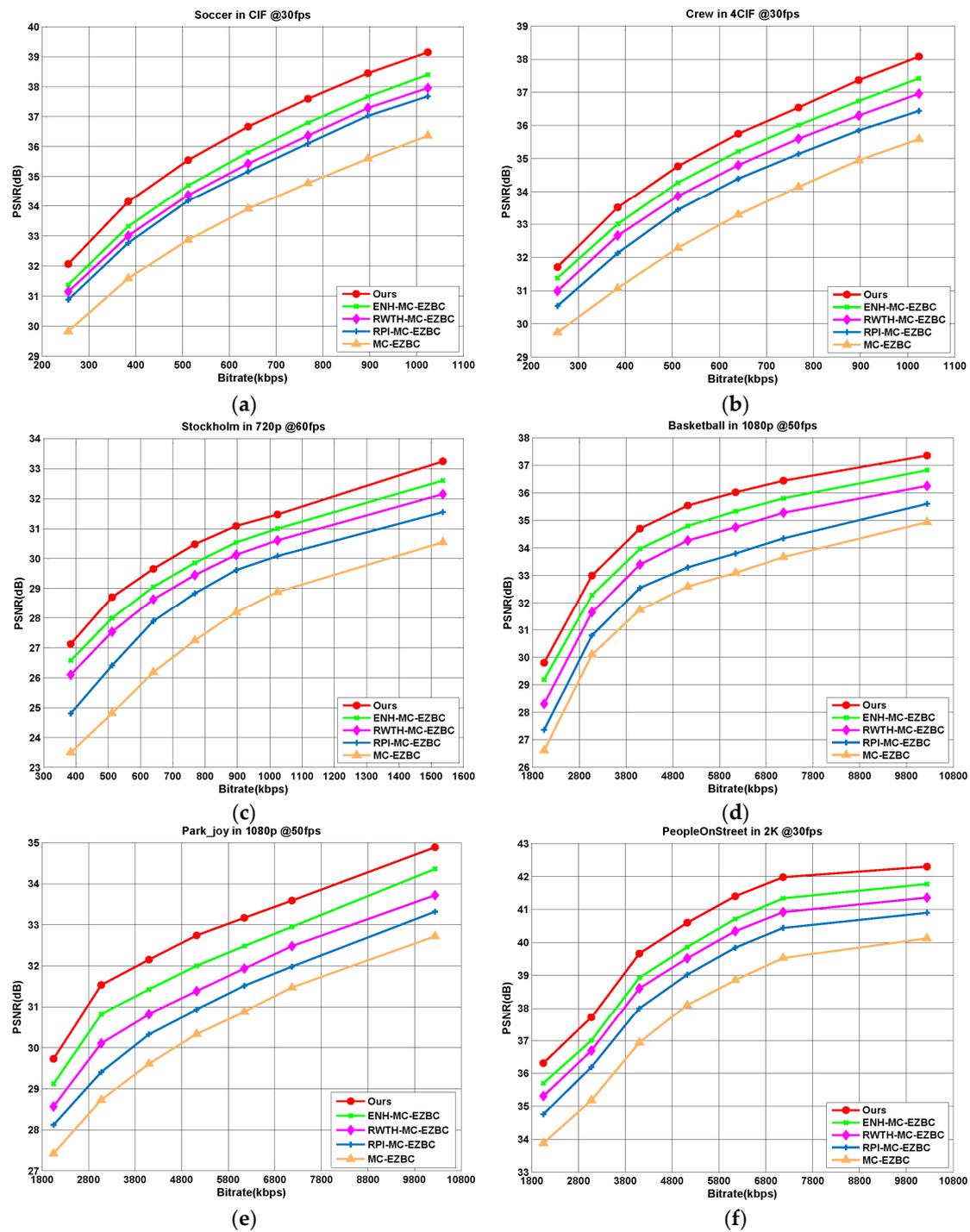


Figure 4. R-D performance comparisons among five different codecs for sequences: (a) Soccer; (b) Crew; (c) Stockholm; (d) Basketball; (e) Park_joy, and (f) PeopleOnStreet.

Average PSNR gains for all the test sequences are also displayed in Figure 5, one can see that the proposed Lagrange multiplier algorithm achieves an average of 0.5–3.57 dB improvement in PSNR results. The PSNR gains are more especially significant for the test sequences with complex texture and/or highly moving objects, such as “Foreman”, “Soccer”, “City”, “Stockholm”, “Basketball”, and “Park_joy” sequences. These test sequences contain abrupt changes over frames in video content characteristics with fast moving objects and highly spatial details and tend to be encoded with various coding types. Thus, for all target bitrates ranges, our method shows the remarkable superiorities of RD performances with higher PSNR gains by performing content-adaptive Lagrange multiplier selection.

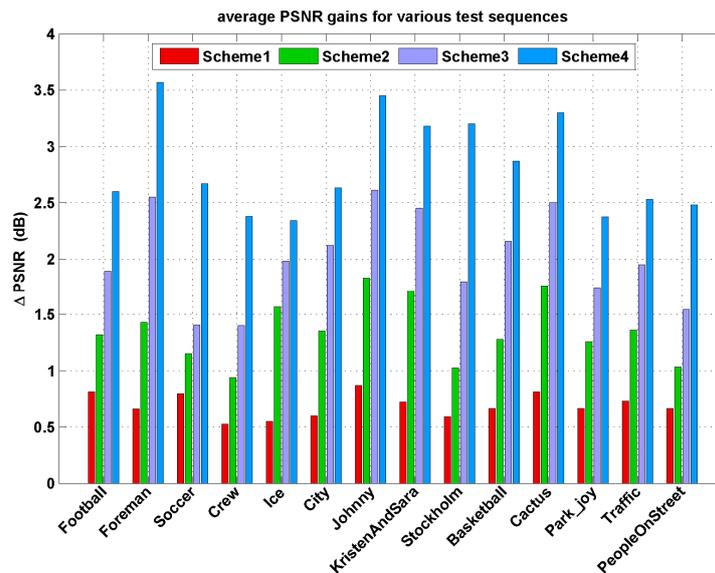


Figure 5. The average PSNR gains obtained for various test sequences.

Figure 6 shows the average PSNR of test video sequences “Soccer” and “Park_joy” at different target bitrates for various Lagrange multiplier values, where the ideal points and the points obtained by the proposed Lagrange multiplier selection are respectively marked. The observation indicates again that the proposed Lagrange multiplier results in the near-ideal rate-distortion performance for 3-D wavelet-based SVC at different target bitrates.

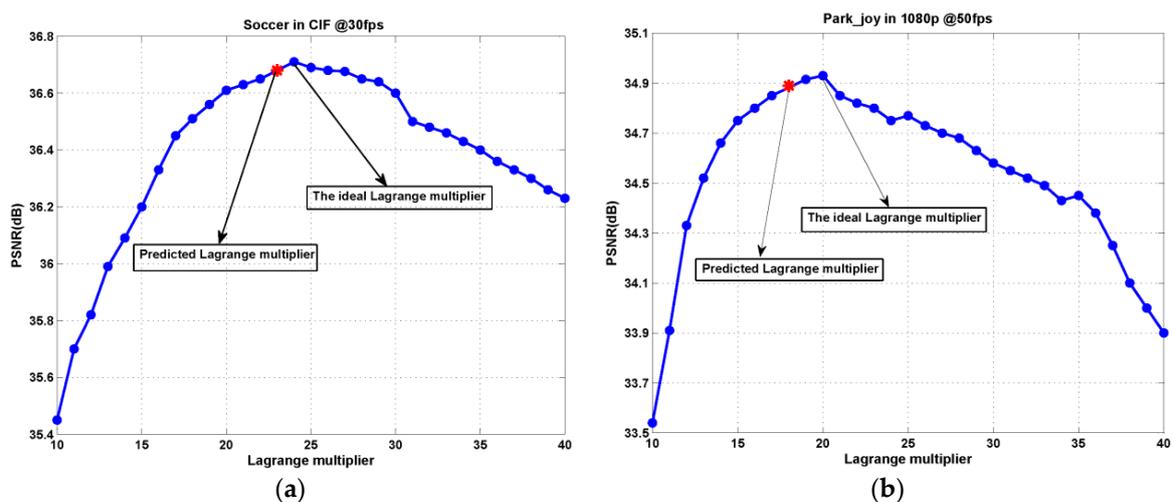


Figure 6. Average PSNR versus Lagrange multiplier at different target bitrates for test video sequences: (a) Soccer (640 kbps) and (b) Park_joy (10240 kbps).

In addition, we also investigate PSNR variations during video reconstruction. Since high fluctuation of frame PSNR values may cause perceptual annoying to viewers, the fluctuation of PSNR values is one of the vital factors for video coding applications. The standard deviation of PSNR (PSNR STD) is utilized to measure for the smoothness of video quality. The smaller PSNR STD value result in the smoother PSNR variation and hence more consistent video quality over the video frames. Figure 7 illustrates the average PSNR STD values over different target bitrates for the proposed algorithm and the other four methods. Compared to the four methods, the proposed algorithm achieves more stable visual quality with substantially smaller PSNR fluctuations as depicted in Figure 7. From this figure, we can see that the proposed Lagrange multiplier selection algorithm reduces the PSNR standard deviations of all frames by up to 3.58 with respect to RPI-MC-EZBC. As shown in Figure 7, our method generates the minimum PSNR fluctuation over the entire sequence than the other three methods. Hence, we reach the conclusion that the proposed method can be beneficial for controlling the fluctuations of qualities in the reconstructed video and produce more stable visual quality than others.

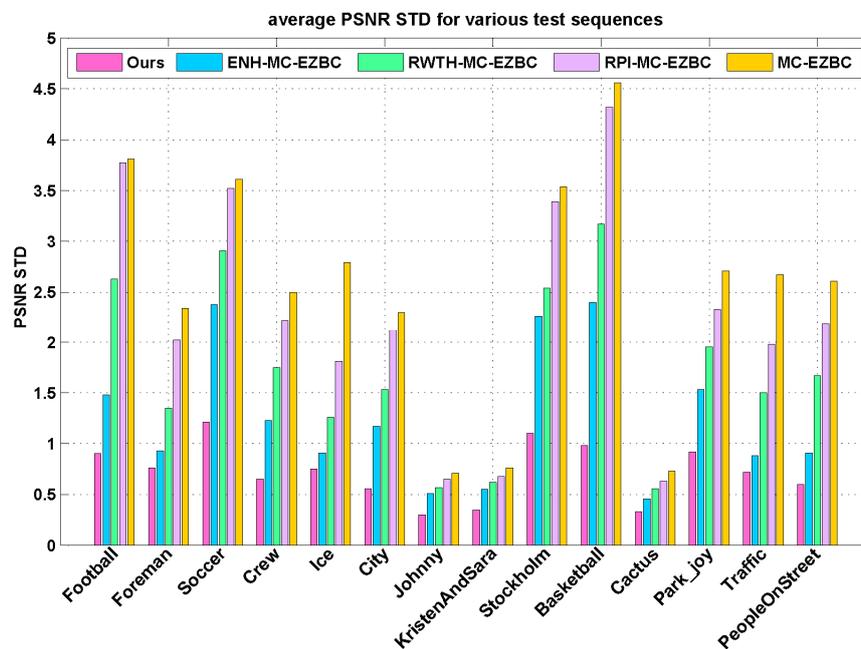


Figure 7. The average standard deviations of PSNR for various test sequences.

4.3.2. Comparison of Subjective Performance

To obtain the subjective inspection of the reconstructed frames, we show the 8th reconstructed frame of the “City” sequence at 896 kbps in Figure 8. From this figure, we can see that the visual quality of the reconstructed frame by the proposed method is conspicuously better than those by the other four methods. It can be plainly discerned that the frame processed by our codec presents less blocking artifacts in the homogeneous regions, better preserved textures, and sharper appearance than other codecs. In particular, it is worth noting that the regions with high spatial details (those enclosed by red rectangles on buildings) are well preserved by the proposed algorithm whereas they are not very clear by other reference codecs.

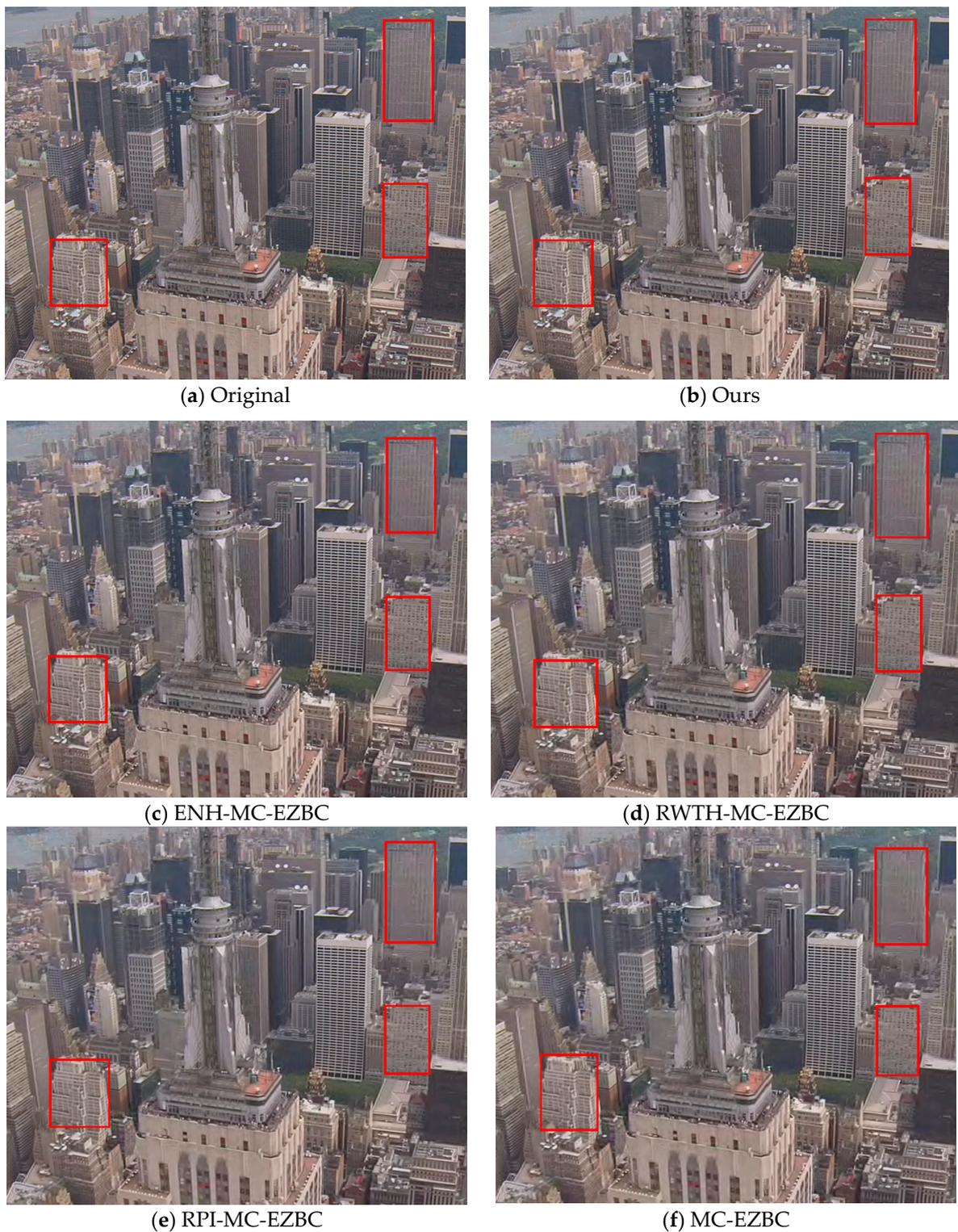


Figure 8. Subjective visual quality comparisons of the 8th reconstructed frame of “City” sequence at 896 kbps.

4.3.3. Comparison of Computational Complexity

To measure the computational complexity, we define the encoding speed as the number of frames which can be encoded in one second on the hardware platform with processor Intel Core i5 4-core CPU 3.10GHz. The computational complexity is considered as inverse value of the encoding

speed, which is measured without any use of assemblers, threads, or other program optimization techniques [6,7,15,28]. Table 4 demonstrates the encoding speed for the proposed algorithm, ENH-MC-EZBC, RWTH-MC-EZBC, RPI-MC-EZBC, and MC-EZBC. As shown in Table 4, the encoding speed of the proposed algorithm is about 1.14, 3.65, 4.65 times faster than RWTH-MC-EZBC, RPI-MC-EZBC, and MC-EZBC, respectively. The reason is that for the sequences with strong inter-frame dependencies, more bits are allocated to the reference frames which lead to better prediction results. Thus, only small residual signals need to be processed in the following encoding steps, which further result in computational complexity reductions. However, the encoding speed of ENH-MC-EZBC is about 1.11 times faster than our algorithm. This is mainly due to the additional operations on the statistics for video content characteristics. In general, the experimental results demonstrate that the computing overhead brought by the proposed Lagrange multiplier selection algorithm can be negligible when compared to the other four methods. Although our algorithm encodes at average 3.07 frames per second, it is insufficient for real-time video processing applications.

Since the ENH-MC-EZBC reference software is non-optimized C++ implementation, development of an efficient low-complexity 3-D wavelet-based scalable video coding scheme is an important practical problem, which is necessary to be considered in our future work.

Table 4. Encoding speed comparison results.

Resolution	Encoding Speed				
	Ours	ENH-MC-EZBC	RWTH-MC-EZBC	RPI-MC-EZBC	MC-EZBC
CIF	5.31	5.82	4.87	1.52	1.21
4CIF	5.16	5.55	4.52	1.29	1.05
720p	3.05	3.57	2.71	0.87	0.64
1080p	1.09	1.34	0.83	0.35	0.29
2K	0.72	0.79	0.55	0.16	0.12
Average	3.07	3.41	2.70	0.84	0.66

5. Conclusions

In this paper, we present an efficient content adaptive Lagrange multiplier selection algorithm for RDO in 3-D wavelet-based SVC. The wavelet filter types, subband coupling in the MCTF process, and temporal subband content characteristics have been incorporated into our algorithm to select the Lagrange multiplier adaptively. The simulation results demonstrate that the proposed algorithm turns out to be much better than the reference methods in terms of both accuracy and effectiveness.

In a future work, we are going to extend our algorithm to other scalabilities, not merely quality scalability. The overall video quality can be further improved by employing human visual system-based perceptual features. In addition, development of a scalable low-complexity video codec based on 3-D DWT is our main concern as well. Therefore, we will experiment in these directions to obtain a more compelling result.

Acknowledgments: This work is supported by the National Science Foundation of China (NSFC) Projects No. 61173110 and No. 6170011056. The authors would like to thank the reviewers for their helpful and detailed suggestions.

Author Contributions: Authors are listed in order of their degree of involvement in the work, with the most active contributors listed first. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, H.; Kao, M.P.; Nguyen, T.Q. Bidirectional scalable motion for scalable video coding. *IEEE Trans. Image Process.* **2010**, *19*, 3059–3064. [[CrossRef](#)] [[PubMed](#)]

2. Wu, Y.S.; Yeh, C.H.; Tseng, W.Y.; Chang, M.K. Scalable video streaming transmission over cooperative communication networks based on frame significance analysis. In Proceedings of the 2012 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC), Hong Kong, China, 12–15 August 2012; pp. 274–279.
3. Chen, D. New Bilinear Model for Inter-Frame Video Compression. Ph.D. Thesis, Rensselaer Polytechnic Institute, New York, NY, USA, 2014.
4. Condoluci, M.; Araniti, G.; Molinaro, A.; Iera, A. Multicast resource allocation enhanced by channel state feedbacks for multiple scalable video coding streams in LTE networks. *IEEE Trans. Veh. Technol.* **2016**, *65*, 2907–2921. [[CrossRef](#)]
5. Omidyeganeh, M.; Ghaemmaghani, S.; Shirmohammadi, S. Application of 3D-wavelet statistics to video analysis. *Multimed. Tools Appl.* **2013**, *65*, 441–465. [[CrossRef](#)]
6. Belyaev, E.; Egiazarian, K.; Gabbouj, M. A low-complexity bit-plane entropy coding and rate control for 3-D DWT based video coding. *IEEE Trans. Multimedia* **2013**, *15*, 1786–1799. [[CrossRef](#)]
7. Belyaev, E.; Egiazarian, K.; Gabbouj, M.; Liu, K. A Low-complexity joint source-channel video coding for 3-D DWT codec. *J. Commun.* **2013**, *8*, 893–901. [[CrossRef](#)]
8. Das, A.; Hazra, A.; Banerjee, S. An efficient architecture for 3-D discrete wavelet transform. *IEEE Trans. Circuits Syst. Video Technol.* **2010**, *20*, 286–296. [[CrossRef](#)]
9. Jin, X.S.; Lin, M.L.; Zhao, Z.J.; Chen, M.; Fan, Z.P. Bitrate allocation for 3-D wavelet video coder. In Proceedings of the 2011 IEEE International Conference on Advanced Computer Control (ICACC), Harbin, China, 18–20 January 2011; pp. 261–264.
10. Belyaev, E.; Forchhammer, S.; Codreanu, M. Error concealment for 3-D DWT based video codec using iterative thresholding. *IEEE Commun. Lett.* **2017**, *21*, 1731–1734. [[CrossRef](#)]
11. Fradj, B.B.; Zaid, A.O. Scalable video coding using motion-compensated temporal filtering and intra-band wavelet based compression. *Multimed. Tools Appl.* **2014**, *69*, 1089–1109. [[CrossRef](#)]
12. Wu, Y.; Woods, J.W. Aliasing reduction via frequency roll-off for scalable image/video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 48–58.
13. Secker, A.; Taubman, D. Highly scalable video compression using a lifting-based 3D wavelet transform with deformable mesh motion compensation. In Proceedings of the 2002 IEEE International Conference on Image Processing (ICIP), Rochester, NY, USA, 22–25 September 2002; pp. 749–752.
14. Xiong, R.; Xu, J.; Wu, F.; Li, S. Barbell-lifting based 3-D wavelet coding scheme. *IEEE Trans. Circuits Syst. Video Technol.* **2007**, *17*, 1256–1269. [[CrossRef](#)]
15. Belyaev, E.; Vinel, A.; Surak, A.; Gabbouj, M.; Jonsson, M.; Egiazarian, K. Robust vehicle-to-infrastructure video transmission for road surveillance applications. *IEEE Trans. Veh. Technol.* **2015**, *64*, 2991–3003. [[CrossRef](#)]
16. Li, S.; Zhu, C.; Gao, Y.; Zhou, Y.; Dufaux, F. Lagrangian multiplier adaptation for rate-distortion optimization with inter-frame dependency. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 117–129. [[CrossRef](#)]
17. Li, S.; Xu, M.; Wang, Z.; Sun, X. Optimal bit allocation for CTU level rate control in HEVC. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 2409–2424. [[CrossRef](#)]
18. Wang, M.; Ngan, K.N.; Li, H. Low-delay rate control for consistent quality using distortion-based Lagrange multiplier. *IEEE Trans. Image Process.* **2016**, *25*, 2943–2955. [[CrossRef](#)] [[PubMed](#)]
19. Gong, Y.; Wan, S.; Yang, K.; Wu, H.; Liu, Y. Temporal-layer-motivated lambda domain picture level rate control for random-access configuration in H.265/HEVC. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 2109–2123. [[CrossRef](#)]
20. Gonzalez-de-Suso, J.L.; Jimenez-Moreno, A.; Martine z-Enriquez, E.; Diaz-de-Maria, F. Improved method to select the Lagrange multiplier for rate-distortion based motion estimation in video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2014**, *24*, 452–464. [[CrossRef](#)]
21. Kuo, C.H.; Shih, Y.L.; Yang, S.C. Rate control via adjustment of Lagrange Multiplier for video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 2069–2078. [[CrossRef](#)]
22. Zhang, F.; Bull, D.R. An adaptive Lagrange multiplier determination method for rate-distortion optimization in hybrid video codecs. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 671–675.

23. Zhao, L.; Zhang, X.; Tian, Y.; Wang, R.; Huang, T. A background proportion adaptive Lagrange multiplier selection method surveillance video on HEVC. In Proceedings of the 2013 IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, USA, 15–19 July 2013; pp. 1–6.
24. Sullivan, G.J.; Wiegand, T. Rate-distortion optimization for video compression. *IEEE Signal Process. Mag.* **1998**, *15*, 74–90. [[CrossRef](#)]
25. Li, X.; Oertel, N.; Hutter, A.; Kaup, A. Advanced Lagrange multiplier selection for hybrid video coding. In Proceedings of the 2007 IEEE International Conference on Multimedia and Expo (ICME), Beijing, China, 2–5 July 2007; pp. 364–367.
26. Seo, S.W.; Han, J.K.; Nguyen, T.Q. Rate control scheme for consistent video quality in scalable video codec. *IEEE Trans. Image Process.* **2011**, *20*, 2166–2176. [[PubMed](#)]
27. Hou, J.; Wan, S.; Ma, Z.; Chau, L. Consistent video quality control in scalable video coding using dependent distortion quantization model. *IEEE Trans. Broadcast.* **2013**, *59*, 717–724. [[CrossRef](#)]
28. Song, J.; Yang, F.; Zhou, Y.; Gao, S. Parametric planning model for video quality evaluation of IPTV services combining channel and video characteristics. *IEEE Trans. Multimedia* **2017**, *19*, 1015–1029. [[CrossRef](#)]
29. Ohm, J.R. Three-dimensional subband coding with motion compensation. *IEEE Trans. Image Process.* **1994**, *3*, 559–571. [[CrossRef](#)] [[PubMed](#)]
30. Xiong, R.; Xu, J.; Wu, F.; Li, S.; Zhang, Y. Optimal subband rate allocation for spatial scalability in 3D wavelet video coding with motion aligned temporal filtering. In Proceedings of the 2005 IEEE International Conference on Visual Communications and Image Processing (VCIP), Beijing, China, 10–13 November 2005; pp. 1–12.
31. Xiong, R.; Xu, J.; Wu, F.; Li, S.; Zhang, Y. Subband coupling aware rate allocation for spatial scalability in 3-D wavelet video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2007**, *17*, 1311–1324. [[CrossRef](#)]
32. Peng, G.; Hwang, W.; Chen, S. Optimal bit-allocation for wavelet-based scalable video coding. In Proceedings of the 2012 IEEE International Conference on Multimedia and Expo (ICME), Melbourne, Australia, 9–13 July 2012; pp. 663–668.
33. Peng, G.; Hwang, W. Optimal bit-allocation for wavelet scalable video coding with user preference. In *Advanced Video Coding for Next-Generation Multimedia Services*; Ho, Y.S., Ed.; InTech Press: Rijeka, Croatia, 2013; pp. 117–138, ISBN 978-953-51-0929-7.
34. Hsiang, S.T.; Woods, J.W. Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank. *Signal Process. Image Commun.* **2001**, *16*, 705–724. [[CrossRef](#)]
35. Wu, Y.; Hanke, K.; Rusert, T.; Woods, J.W. Enhanced MC-EZBC scalable video coder. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 1432–1436.
36. Ortega, A.; Ramchandran, K. Rate-distortion methods for image and video compression. *IEEE Signal Process. Mag.* **1998**, *15*, 23–50. [[CrossRef](#)]
37. Everett, H. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Oper. Res.* **1963**, *11*, 399–417. [[CrossRef](#)]
38. Ramchandran, K.; Vetterli, M. Best wavelet packet bases in a rate-distortion sense. *IEEE Trans. Image Process.* **1993**, *2*, 160–175. [[CrossRef](#)] [[PubMed](#)]
39. Schuster, G.M.; Katsaggelos, A.K. An optimal quadtree-based motion estimation and motion-compensated interpolation scheme for video compression. *IEEE Trans. Image Process.* **1998**, *7*, 1505–1523. [[CrossRef](#)] [[PubMed](#)]
40. Wiegand, T.; Schwarz, H.; Joch, A.; Kossentini, F.; Sullivan, G.J. Rate-constrained coder control and comparison of video coding standards. *IEEE Trans. Circuits Syst. Video Technol.* **2003**, *13*, 688–703. [[CrossRef](#)]
41. Hsiang, S.T.; Woods, J.W.; Ohm, J.R. Invertible temporal subband/wavelet filter banks with half-pixel-accurate motion compensation. *IEEE Trans. Image Process.* **2004**, *13*, 1018–1028. [[CrossRef](#)] [[PubMed](#)]
42. Tsai, C.Y.; Hang, H.M. A rate-distortion analysis on motion prediction efficiency and mode decision for scalable wavelet video coding. *J. Vis. Commun. Image Represent.* **2010**, *21*, 917–929. [[CrossRef](#)]
43. Sharabayko, M.P.; Ponomarev, O.G. Fast rate estimation for RDO mode decision in HEVC. *Entropy* **2014**, *16*, 6667–6685. [[CrossRef](#)]
44. Liu, Z.G.; Peng, Y.H.; Yang, Y. An adaptive GOP structure selection for haar-like MCTF encoding based on mutual information. *Multimed. Tools Appl.* **2009**, *43*, 25–43. [[CrossRef](#)]

45. Xu, R.; Taubman, D.; Naman, A.T. Motion estimation based on mutual information and adaptive multi-scale thresholding. *IEEE Trans. Image Process.* **2016**, *25*, 1095–1108. [[CrossRef](#)] [[PubMed](#)]
46. Bar-Yosef, Y.; Bistriz, Y. Gaussian mixture models reduction by variational maximum mutual information. *IEEE Trans. Signal Process.* **2015**, *63*, 1557–1569. [[CrossRef](#)]
47. Taubman, D. High performance scalable image compression with EBCOT. *IEEE Trans. Image Process.* **2000**, *9*, 1158–1170. [[CrossRef](#)] [[PubMed](#)]
48. Mallat, S.; Falzon, F. Analysis of low bit rate image transform coding. *IEEE Trans. Signal Process.* **1998**, *46*, 1027–1042. [[CrossRef](#)]
49. Wan, S.; Yang, F.; Izquierdo, E. Lagrange multiplier selection in wavelet-based scalable video coding for quality scalability. *Signal Process. Image Commun.* **2009**, *24*, 730–739. [[CrossRef](#)]
50. Devore, J.; Farnum, N.; Doi, J. *Applied Statistics for Engineers and Scientists*; Cengage Learning Press: Boston, MA, USA, 2013.
51. Anegekuh, L.; Sun, L.; Jammeh, E.; Mkvawa, I.; Ifeachor, E. Content-based video quality prediction for HEVC encoded videos streamed over packet networks. *IEEE Trans. Multimed.* **2015**, *17*, 1323–1334. [[CrossRef](#)]
52. Banitalebi-Dehkordi, A.; Pourazad, M.T.; Nasiopoulos, P. An efficient human visual system based quality metric for 3D video. *Multimed. Tools Appl.* **2016**, *75*, 4187–4215. [[CrossRef](#)]
53. MC-EZBC Video Coder Software Repository. Available online: <http://www.cipr.rpi.edu/research/mcezbc/> (accessed on 30 January 2018).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).