Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2009, Article ID 109057, 11 pages
doi:10.1155/2009/109057

Research Article

Spatial-Aided Low-Delay Wyner-Ziv Video Coding

Bo Wu,[1] Xiangyang Ji,[2] Debin Zhao,[3] and Wen Gao[1,4]

[1] *Digital Media Research Center, Institute of Computing Technology, Chinese Academy of Science, Beijing 100190, China*
[2] *Department of Automation, Tsinghua University, Beijing 100084, China*
[3] *Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China*
[4] *Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China*

Correspondence should be addressed to Debin Zhao, dbzhao@jdl.ac.cn

Received 6 May 2008; Revised 28 October 2008; Accepted 12 January 2009

Recommended by Anthony Vetro

In distributed video coding, the side information (SI) quality plays an important role in Wyner-Ziv (WZ) frame coding. Usually, SI is generated at the decoder by the motion-compensated interpolation (MCI) from the past and future key frames under the assumption that the motion trajectory between the adjacent frames is translational with constant velocity. However, this assumption is not always true and thus, the coding efficiency for WZ coding is often unsatisfactory in video with high and/or irregular motion. This situation becomes more serious in low-delay applications since only motion-compensated extrapolation (MCE) can be applied to yield SI. In this paper, a spatial-aided Wyner-Ziv video coding (WZVC) in low-delay application is proposed. In SA-WZVC, at the encoder, each WZ frame is coded as performed in the existing common Wyner-Ziv video coding scheme and meanwhile, the auxiliary information is also coded with the low-complexity DPCM. At the decoder, for the WZ frame decoding, auxiliary information should be decoded firstly and then SI is generated with the help of this auxiliary information by the spatial-aided motion-compensated extrapolation (SA-MCE). Theoretical analysis proved that when a good tradeoff between the auxiliary information coding and WZ frame coding is achieved, SA-WZVC is able to achieve better rate distortion performance than the conventional MCE-based WZVC without auxiliary information. Experimental results also demonstrate that SA-WZVC can efficiently improve the coding performance of WZVC in low-delay application.

1. Introduction

Recently, the new applications such as wireless video surveillance and wireless sensor network are emerging. In these applications, a light encoder is required because the computation and memory resources on sensors are scarce. Furthermore, in these systems, there are always a high number of encoders and only one or a few decoders. As a result, the conventional hybrid video coding architectures such as H.26x and MPEG-x, are no longer being applicable due to the intrinsic one-to-many application model with one high-complexity encoder and many low-complexity decoders. In theory, distributed source coding (DSC) can provide an ideal solution to address this problem. The Slepian-Wolf theory shows that under certain conditions, even if the correlated sources are encoded separately and decoded jointly, the coding performance can be as good as joint encoding and decoding [1]. Later, Wyner and Ziv extended this theory to the lossy source coding with side information (SI) at the decoder [2], which is more suitable for practical video coding. Many researchers have applied the practical WZ coding techniques in video coding [3–5]. One advantage of WZ coding is that the computational complexity of the encoder is low, such as those schemes proposed in [4, 5]. In these schemes, the motion correlation does not need to be exploited at the encoder and the frames are only compressed by low-complexity channel coding method, such as turbo codes. While at the WZ decoder, the motion estimation with high computational complexity is applied to exploit the temporal correlation in SI generation. Subsequently, the errors between the original information and the SI are corrected by using the received parity bits transmitted from the encoder. Another advantage of WZVC is the robustness since the WZVC system is drift-free due to no motion estimation and motion compensation prediction at the encoder. WZVC system is also deemed one type of

the joint source-channel coding systems [6] since it can be used as a systematic lossy forward error protection method for conventional video coding.

In [3], two typical SI generation approaches are introduced, which are motion-compensated interpolation (MCI) and extrapolation (MCE), respectively. For MCI, SI for the current frame is yielded by performing motion compensation on the adjacent previously and subsequently decoded picture. However, in low-delay application, the temporally subsequent pictures cannot be used as references to generate SI. Therefore, MCE is adopted to generate SI in low-delay application, in which the motion between the decoded frames at time t_2 and time t_1 are estimated and the estimated motion are used to extrapolate the SI at time t. However, the performance of MCE-based low-delay WZVC is often unsatisfactory because motion field cannot be well estimated [3]. In fact, this situation can be improved by the auxiliary information-aided method, in which partial information of the current frame is used as the auxiliary information to help the decoder to improve the accuracy of motion field for MCE. In [7], one frame is partitioned into intra- and WZmacroblocks by a pattern which is similar to H.264/AVC FMO grouping method. The subset of intra-macroblocks is employed as auxiliary information and helps for estimating the SI with temporal concealment method. The auxiliary information-aided method can also be used to improve the quality of SI in the case of MCI. In [5], the quantized DCT domain coefficients named hash bits are performed as the auxiliary information. In [8], a coarse representation of the frame is considered to assist motion estimation at the decoder. For the above auxiliary information-aided WZ coding schemes, significant improvements of performance can always be achieved.

The discrete wavelet transform (DWT) are highly desirable for video coding due to their intrinsic multiresolution structure and energy compaction property. For hybrid video coding, DWT has been applied in many state-of-art coding schemes to obtain the spatial scalable functionality, such as [9, 10]. Moreover, in DVC paradigm, the DWT also has been widely used. In [11], the author explored the high-order statistical correlation among the transform coefficients by using DWT and SPHIT algorithms. In [12], hyperspectral images from neighboring frequency bands are closely correlated. The authors propose a prediction model based on linear prediction techniques. Under the model, the correlation among bit-planes from neighboring DWT bands is exploited. In [13], the authors used the shift-invariant redundant discrete wavelet transform (RDWT) reference frames for finding matching blocks to overcome the inefficiency of motion estimation in critically sampled wavelet domain. In [14], the authors proposed a context correlation model between the source and its SI in the wavelet transform domain. Compared to RDWT domain motion estimation and motion compensation, spatial domain motion estimation and motion compensation are usually able to yield better prediction efficiency [9].

To improve low-delay WZ coding, this paper proposes a spatial-aided WZ video coding scheme. The spatial DWT, which inherently supports spatial scalability, is used to generate auxiliary information. At the encoder, one WZ frame is decomposed by a spatial 2D-DWT first and its low-pass subband is used as the auxiliary information. First, the auxiliary information is encoded by DPCM coding method and thus, the partial correlation among adjacent auxiliary information can be removed. Then, the whole-frame is encoded by DCT domain Wyner-Ziv encoder. At the decoder, auxiliary information should be decoded firstly. Then SI is generated by the SA-MCE algorithm in which motion field for generating SI is achieved by performing motion estimation on the spatial auxiliary information and the low-pass subband of previously decoded frames in spatial domain. With the help of the auxiliary information, more precise motion field can be obtained. Hence, the spatial-aided Wyner-Ziv video coding (SA-WZVC) approach is able to achieve a better rate distortion performance against the conventional MCE-based WZVC without auxiliary information. In addition, due to the inherent decomposition structure of wavelet transform, the scalability can be achieved easily.

The remainder of this paper is organized as follows: Section 2 describes the proposed scheme in detail. Section 3 analyzes the rate distortion performance of the proposed spatial-aided WZ coding method theoretically and compares it with the conventional MCE-based low-delay WZ coding. By using the theoretical model, some numerical results are presented. In Section 4 simulation results are given.

2. Spatial-Aided Low-Delay Wyner-Ziv Video Coding

2.1. Spatial-Aided Low-Delay Wyner-Ziv Video Coding Scheme. As shown in Figure 1, the framework of the spatial-aided low-delay WZ coding is similar to the framework presented in [4]. The key frames of the video sequence are compressed using a conventional intra-frame codec. The remaining frames, namely WZ frames, are encoded by spatial-aided low-delay WZ codec. At the encoder, the auxiliary information generation module is applied to the original WZ frames. The generated spatial auxiliary information is encoded by DPCM coding method, while the whole WZ frame is encoded by DCT transform domain Wyner-Ziv video coding (WZVC) as proposed in [3]. At the decoder, the spatial auxiliary information is decoded first. Subsequently, with the help of the decoded spatial auxiliary information, the spatial-aided motion-compensated extrapolation- (SA-MCE-) based SI generation algorithm is performed. At last, the WZ frame is decoded by the DCT domain WZ decoder. The detail of each part in the system is described as follows.

2.2. Spatial Auxiliary Information Coding. There are many methods for the auxiliary information generation such as [5, 7, 8]. Considering the energy compaction characteristics of DWT, DWT is adopted as a tool to generate the auxiliary information. At the encoder, for each WZ frame, one level 2D-DWT with biorthogonal 9/7 filter is applied to decompose the original frame and the low-low- (LL-) pass subband of current frame is used as spatial auxiliary

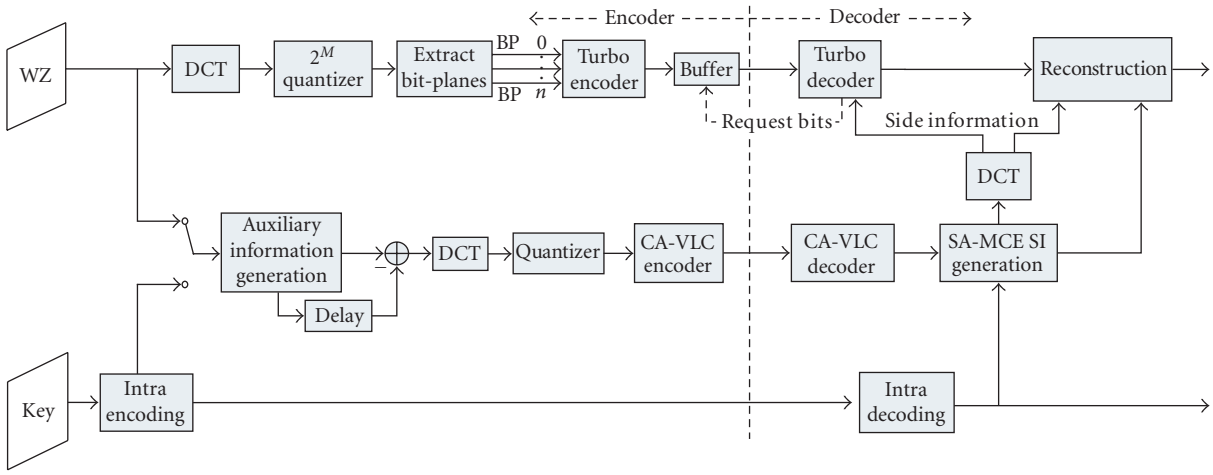

Figure 1: Framework of spatial-aided low-delay WZ codec.

information. As a result, the resolution of the spatial auxiliary information is a quarter of the original frame. To reduce the temporal redundancy, DPCM is performed between the adjacent LL subbands to encode the LL subband. For DPCM coding, the difference between the current LL subband and its previously reconstructed reference frame is calculated. Then the residues are DCT transformed and quantized by a quantizer. Finally, the quantized coefficients are encoded by a CA-VLC entropy encoder used in H.264/AVC. If the reference frame is a key frame, the LL subband of full-resolution reconstructed intra-frame needs to be yielded by DWT to form the reference frame for DPCM coding.

2.3. Wyner-Ziv Frame Coding. At the encoder, the whole WZ frame is encoded by DCT transform domain WZ coding [3]. First, a block-wise DCT is applied to the whole WZ frame and the statistical dependencies within a frame are exploited. The transform coefficients are grouped together to form the coefficient bands. Then for each band, different M-level uniform scalar quantizers are applied. Next, the bit-planes are extracted and each bit-plane is organized to fixed length binary codewords. Each codeword is sent to the Slepian-Wolf (SW) encoder as input and the output is the parity bits. The SW coder is implemented using a rate-compatible punctured turbo code (RCPT). Then, these parity bits are punctured into different blocks and stored in a buffer. The blocks of parity bits, which are also called as WZ bits, are successively transmitted to decoder upon request.

At the decoder, the spatial auxiliary information of current WZ frame is decoded first. Then, the SI of whole WZ frame is generated with the help of the auxiliary information by an SA-MCE method which is presented in Section 2.5. Subsequently, DCT is applied on the generated full-resolution SI and the coefficients in each DCT block are extracted into different subbands corresponding to the DCT bands partition patterns. The DCT coefficient Y_i of SI at the ith position in current subband is used for the bit-plane probabilities evaluation. This means that for every original coefficient X_i the value of Y_i is used to evaluate the probability of every bit of X_i being 1 or 0. The detailed description about the probability evaluation and correlation model being used is introduced in the next subsection.

2.4. Correlation Model. As the turbo decoder obtains the side information, a priori probability of current decoding bit-planes should be calculated first. According to simulation results, the probability distribution of the difference between the source and its SI conforms to a Laplacian model and thus, the Laplacian model is taken as the probability density function for calculating the a priori probability. To estimate the values of the jth bit of X_i being 0 or 1, the probability can be calculated as

$$p(\overline{b}_i^j \mid y_i, s_i, b_i^0, \ldots, b_i^{j-1}) = \frac{\alpha}{2} e^{-\alpha|d|} \tag{1}$$

with

$$\begin{aligned} d &= a \cdot (Z_i + \mathit{offset}) - y_i \\ &= a\{(b_i^0 \cdot 2^m + \cdots + b_i^{j-1} \cdot 2^{m-j+1} + \overline{b}_i^j \cdot 2^{m-j}) \\ &\qquad + 2^{m-j-1}\} - y_i. \end{aligned} \tag{2}$$

Let b_i^j denote the jth bit-plane at the position i in current subband and its estimation is $\overline{b}_i^j$. However, $\{b_i^0, \ldots, b_i^{j-1}\}$ are those previously decoded bits and b_i^0 is the most significant bit. In (1), S_i is the sign bit. If the coefficient X_i is positive, S_i equals 0; otherwise S_i equals 1. For each coefficient band, different standard deviation of Laplacian model $1/\alpha$ is adopted. The value of $1/\alpha$ is determined by offline training.

In (2), Z_i represents the integer number that has the jth bit $\overline{b}_i^j$ and those previously more significant bits $\{b_i^0, \ldots, b_i^{j-1}\}$. *Offset* is an estimated value used to compensate the lower part of Z_i. If X_i is partitioned into m bins, offset equals 2^{m-j-1}. a is used to adjust the sign of the value $(Z_i + \mathit{offset})$, which is defined as

$$a = \begin{cases} 1 & s_i = 0, \\ -1 & s_i = 1. \end{cases} \tag{3}$$

According to (1), (2), and (3), the transition probability on branches in trellis of turbo code can be obtained. When the decoder receives the parity bits, the trellis graph is traversed for several times. If the bit-error rate (BER) of current bit-plane converges to an acceptable value, the request for parity bits stopped and the current bit-plane is successively decoded. Otherwise, more parity bits are required. After the current bit-plane is decoded, it is used in calculating the a priori probability of next bit-plane as defined in (1).

2.5. SA-MCE-Based Side Information Generation. Motion-compensated extrapolation is a general method in low-delay WZ coding schemes. For the MCE method, as shown in [3], the motion between the decoded frames at time t_1 and time t_2 are estimated and the estimated motion is used to extrapolate the SI at time t. However, due to the absence of information of current frame, the MCE method is not very effective. Therefore, spatial auxiliary information-aided MCE method is adopted in this paper.

The proposed SA-MCE SI generation scheme is depicted as Figure 2. The detailed procedure is as follows. In order to obtain the motion information for motion compensation at high resolution, the low-resolution auxiliary information needs to be upsampled first. Subsequently, motion search can be performed on current upsampled low-resolution frame and previous upsampled low-resolution frames (LL), or on current upsampled low-resolution frame and previous reconstructed high resolution frames (L-H). Due to the lack of high-pass subband, those upsampled low-resolution frames suffer from the artifacts, such as blending, aliasing, and tiling. As shown in [15] the artifacts in the upsampled low-resolution frames (L) can disturb block matching when compared to the blocks in the high-quality reference frame (H). The previously upsampled low-resolution frames have the same artifacts, so the effect of artifacts could be nullified by the similar blocking artifacts. Therefore, it is necessary to perform DWT and IDWT to obtain the upsampled frame of the LL band, even for the case of the previous frame being key frame. The inverse DWT transform IDWT_{L_0} operator is used to upsample the LL subband and it is defined as follows:

$$\Delta\hat{X}_{LL(t)} = \mathrm{IDWT}_{L\_0}\{\hat{X}_{LL(t)}, 0\}, \tag{4}$$

where $\hat{X}_{LL(t)}$ is the LL subband at time t and $\Delta\hat{X}_{LL(t)}$ is the upsampled LL frame at time t. IDWT_{L_0} operator is an inverse DWT in which the LL subband is $\hat{X}_{LL(t)}$ and the high-pass sub-bands are all set to zeros. Secondly, the motion estimation is performed between the upsampled spatial auxiliary information $\Delta\hat{X}^w_{LL(t)}$ and its reference $\Delta\hat{X}^r_{LL(t-1)}$. The reference could be an upsampled LL band of a reconstructed key frame or the upsampled LL band of a reconstructed WZ frame.

In this work, the MVs between the upsampled low-resolution frames are directly used for full-resolution MCE. The previously reconstructed full-resolution frame (either key frame or WZ frame) is used as the reference frame for MCE:

$$\tilde{X}^w_{F(t)}(p) = \hat{X}^r_{F(t-1)}(p + mv), \tag{5}$$

where $\hat{X}^r_{F(t-1)}$ denotes the reconstructed full-resolution frame at time (t_1) and $\tilde{X}^w_{F(t)}$ denotes the motion-compensated full-resolution frame at time t. Because of the interband correlation of DWT transformed coefficients, the high-pass subbands prediction of current WZ frame are also obtained through the motion compensation. Consequently, a full-resolution prediction signal of current WZ frame $\tilde{X}^w_{F(t)}$ is generated by (5).

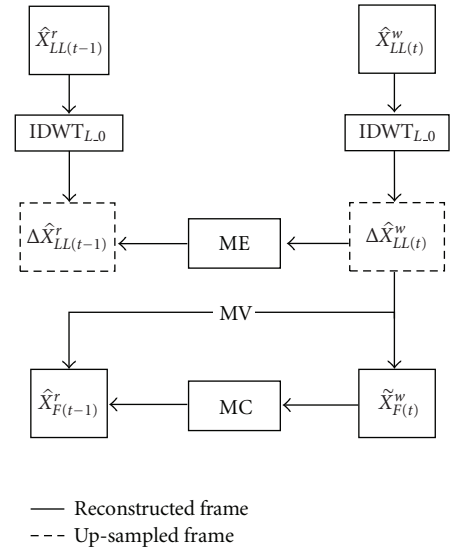

Figure 2: SA-MCE method.

From the numerical results of rate distortion analysis, it can be found that when the quality of auxiliary information is improved adequately, the performance of WZVC can be enhanced. Hence, more bits are allocated to the auxiliary information coding than the WZ frame coding which induces the quality of DPCM-coded LL-band to be high. By means of statistic, it is found that the objective quality of DPCM-coded LL band is better than the LL-band of the extrapolated prediction $\tilde{X}^w_{F(t)}$ in most cases. So the DPCM coded LL subband is substituted for the LL band of the full-resolution prediction $\tilde{X}^w_{F(t)}$ by IDWT operation. The refined SI $Y^w_{F(t)}$ is calculated by

$$Y^w_{F(t)} = \mathrm{IDWT}\{\hat{X}^w_{LL(t)}, \tilde{X}^w_{H(t)}\}, \tag{6}$$

where $\hat{X}^w_{LL(t)}$ is the DPCM-coded LL subbands of WZ frame at time t. Also, $\tilde{X}^w_{H(t)}$ represents three high-pass subbands of $\tilde{X}^w_{F(t)}$ and it is obtained by DWT operation. At last, the side information $Y^w_{F(t)}$ used for WZ decoding is generated.

3. Rate Distortion Analysis

3.1. Background. In conventional hybrid video coding schemes, the motion estimation can be performed at the

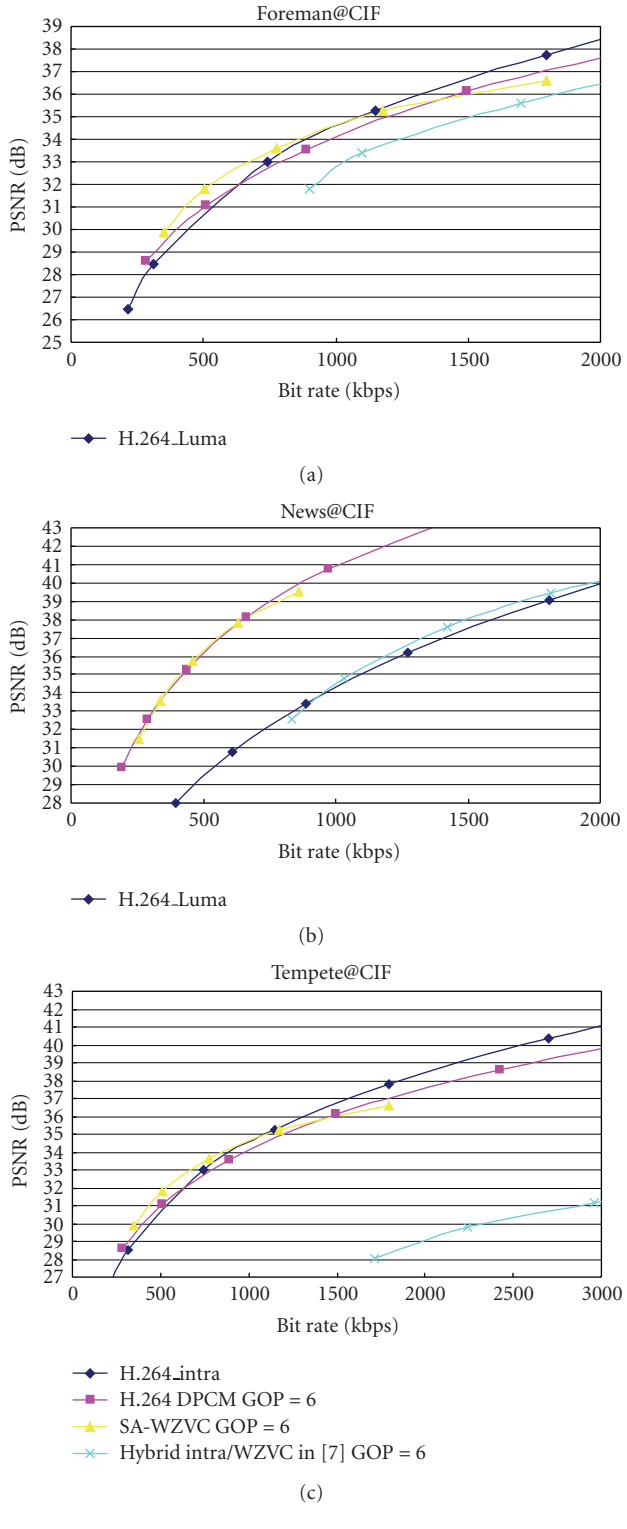

Figure 3: Overall RD performance comparison.

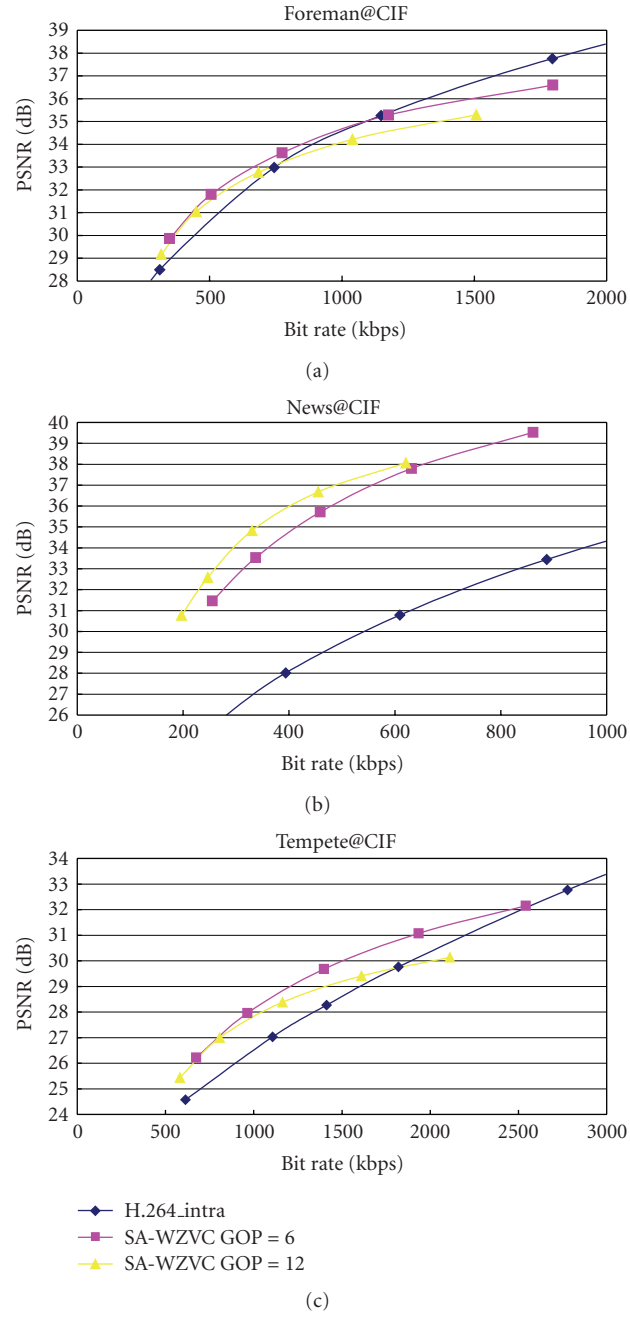

Figure 4: Overall RD performance comparison for varying GOP sizes.

encoder and the accuracy of motion estimation is assumed to be only related to the finite precision used to present the motion vectors. In MCE-based WZVC scheme, the motion estimation is performed at the decoder. Since the current frame is unavailable at the decoder, motion estimation is performed between two previously reconstructed reference frames and the obtained MVs are used to extrapolate the SI of current frame. The MVs between two previous frames do not exactly conform to the MVs between the current frame and its previous reference frame, when the motion trajectory among the adjacent frames is not translational with constant velocity. Therefore, the quality of the side information may not be satisfactory. In our spatial-aided WZVC scheme, the reduced-resolution spatial information is encoded and transmitted to decoder side. The underlying idea is that motion estimation at the decoder has an access to spatial auxiliary information, so the partial description

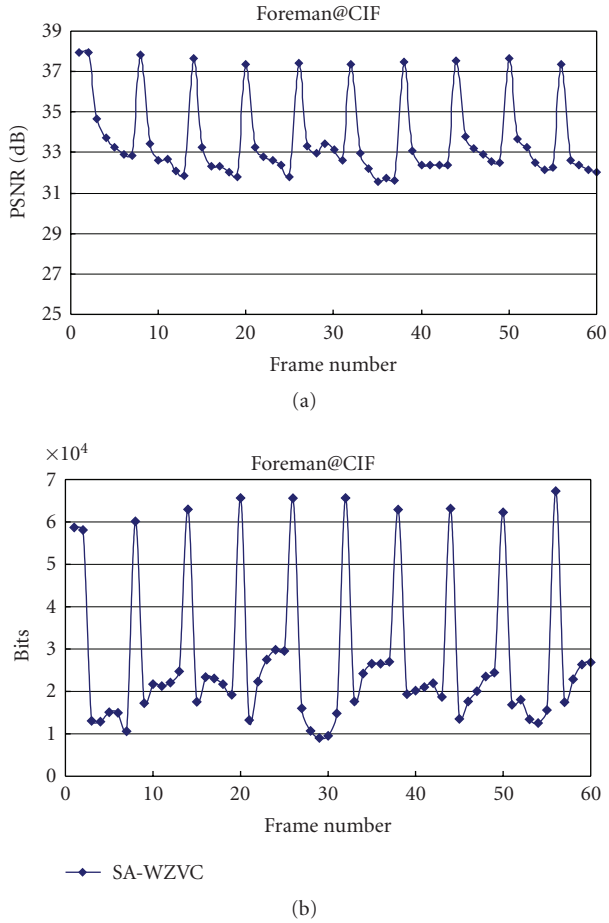

Figure 5: PSNR and Bit-rate trace for SA-WZVC (GOP = 6).

of the current frame may help in obtaining a more accurate estimate of the motion model.

In this work, signal power spectrum and Fourier analysis tools are used to analyze SA-WZVC. The tools are widely used in rate distortion analyzing of hybrid video coding schemes. The rate distortion performance for the conventional MCP-based video coding is analyzed in [16]. Then, the fractional pixel motion search, the long-term motion search, and the multi-hypothesis are studied in [17–20] respectively. Recently, the signal power spectrum methods are also introduced in the Wyner-Ziv coding. In [21], the authors presented a theoretical rate distortion model to examine the WZVC performance and compare it with the conventional motion-compensated prediction-(MCP-) based video coding. The theoretical results show that although WZVC can achieve as much as 6 dB gain in PSNR over the conventional video coding without motion search, it still falls in 6 dB or more in terms of PSNR behind the best MCP-based inter-frame video coding schemes. In [22], the authors studied the theoretical rate distortion model for auxiliary hash-based WZVC scheme. In this scheme, the hash is the high-pass coefficients of DCT transform and this hash is used to perform motion estimation at the decoder. It proves that at high rates, hash-based motion modeling can virtually achieve the same coding efficiency as motion-compensated predictive coding. However, at medium or low rates, a significant coding loss is observed. In this work, these theoretical analysis tools are extended to investigate the rate distortion performance of our spatial-aided low-delay WZVC scheme. During our analysis, some ideas and the theoretical tools are borrowed from the above works and these discussions are meaningful since the optimal generation and coding methods for auxiliary information have not been fully exploited yet.

3.2. Rate Distortion Analysis of Auxiliary Information-Aided Wyner-Ziv Coding. In the following discussions, a rate difference model of SA-WZVC scheme versus the conventional MCE-based WZVC scheme is established. This rate difference model relates the accuracy of the motion model to the power spectral density (PSD) of quantization noise signal. Furthermore, the numerical result of the theoretical model is presented and the result demonstrate that for SA-WZVC scheme, a rate savings can be achieved compared with the conventional MCE-based WZVC when a good trade off between the auxiliary information coding and WZ coding is achieved.

3.2.1. Rate Distortion Analysis. The prediction residual $e(t)$ is

$$e(t) = s(t) - s'(t), \tag{7}$$

where $s(t)$ denotes the input source and $s'(t)$ denotes the MCP frame for the conventional video coding or SI for WZVC.

According to [16, 19], the power spectrum of the prediction residual $\Phi_{ee}(\omega)$ is expressed as

$$\Phi_{ee}(\omega) = 2\Phi_{SS}(\omega)\left(1 - e^{-(1/2)\omega^T \omega \sigma_\Delta^2}\right) + \theta, \tag{8}$$

where $\Phi_{SS}(\omega)$ is the spatial power spectral density (PSD) of the original frame $s(t)$. Also, $\Delta = (\Delta_x, \Delta_y)$ is the motion vector error, that is, the difference between the used motion vectors (MVs) and the true MVs. Finally, θ is the noise term introduced by quantization step.

If it is only considered the prediction inaccuracy introduced by either SA-MCE or MCE, the assumption $\Phi_{SS}(\omega) \gg \theta$ can be made according to [16]. So the difference in rate between intra-frame coding of prediction error e and intra-frame coding of s is obtained as follows according to [16] or [19]:

$$\begin{aligned} \Delta R_{\mathrm{MC}} &= \frac{1}{8\pi^2} \int_{\omega_x} \int_{\omega_y} \log_2 \frac{\Phi_{ee}}{\Phi_{ss}} d\omega \\ &= \frac{1}{8\pi^2} \int_{\omega_x} \int_{\omega_y} \log_2 \left(2\left(1 - e^{-(1/2)\omega^T \omega \sigma_\Delta^2}\right)\right) d\omega. \end{aligned} \tag{9}$$

Hence, the rate difference between the SA-WZVC using MVs (d_x'', d_y'') and MCE-based WZVC using MVs (d_x', d_y') can be yielded by

$$\begin{aligned}\Delta R_f &= \Delta R_{\text{SA-MCE}} - \Delta R_{\text{MCE}} \\ &= \frac{1}{8\pi^2}\int_{\omega_x}\int_{\omega_y} \log_2\left(2\left(1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d2}}\right)\right)d\omega \\ &\quad - \frac{1}{8\pi^2}\int_{\omega_x}\int_{\omega_y} \log_2\left(2\left(1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d1}}\right)\right)d\omega \\ &= \frac{1}{8\pi^2}\int_{\omega_x}\int_{\omega_y} \log_2\left(\frac{1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d2}}}{1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d1}}}\right)d\omega,\end{aligned} \tag{10}$$

where

$$\Delta d1 = (d_x, d_y) - (d_x', d_y') \tag{11}$$

in which (d_x, d_y) denotes the true MVs and (d_x', d_y') denotes the MVs obtained by the MCE algorithm, respectively:

$$\Delta d2 = (d_x, d_y) - (d_x'', d_y''), \tag{12}$$

and (d_x'', d_y'') indicates the MVs obtained by the SA-MCE algorithm.

In our scheme, the spatial auxiliary information $s_l(t)$ is coded by DPCM method. The prediction residual is denoted as $\widetilde{e}_l(t)$. For the DPCM coding of spatial auxiliary information, the R(D) function is

$$R_l(\widetilde{\theta}) = \frac{1}{8\pi^2}\int_{\omega_x}\int_{\omega_y} \log_2\frac{\Phi_{\widetilde{e}_l\widetilde{e}_l}(\omega)}{\widetilde{\theta}}d\omega, \tag{13}$$

where the PSD of the $\widetilde{e}_l(t)$ is

$$\Phi_{\widetilde{e}_l\widetilde{e}_l}(\omega) = 2\Phi_{S_lS_l}(\omega)\left[1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta\text{MV1}}}\right] + \widetilde{\theta}, \quad \omega \in \omega_l, \tag{14}$$

where $\Phi_{S_lS_l}(\omega)$ are the PSD of spatial auxiliary information. Since the MVs of DPCM coding is $(0, 0)$, the motion vector error is equals to the true MVs:

$$\Delta\text{MV1} = (\widetilde{d}_x, \widetilde{d}_y) - (0, 0). \tag{15}$$

Equation (13) is the R(D) function of the spatial auxiliary information coding. The rate difference which takes the spatial correlation of the prediction error $\widetilde{e}_l(t)$ and the original signal $s_l(t)$ into account is widely used to measure the bit-rate reduction. It represents the maximum bit-rate reduction possible by optimum encoding of the prediction error, compared to optimum intra-frame encoding of the signal for the same mean-squared reconstruction error [19]. To obtain an upper bound of rate reduction, the rate difference is measured by comparing the prediction error of auxiliary information $\widetilde{e}_l(t)$ with the prediction error of low-pass subband $\widehat{e}_l(t)$ whose full-resolution frame is encoded with MCE-based WZVC method.

For the MCE-based WZVC, the prediction error of the low-pass subband can be expressed as $\widetilde{e}_l(t)$. The R(D) function of the low-pass subband coding can be expressed as

$$R_l(\widehat{\theta}) = \frac{1}{8\pi^2}\int_{\omega_x}\int_{\omega_y} \log_2\frac{\Phi_{\widehat{e}_l\widehat{e}_l}(\omega)}{\widehat{\theta}}d\omega, \tag{16}$$

where $\widehat{\theta}$ is the PSD of quantization error introduced into the low-pass subband. The MVs (d_x', d_y') in (11) can be taken as the subpixel accuracy MVs of the low-pass subband. To coincide with the motion compensation of the low-pass subband coding, these subpixel accuracy MVs can be reduced to integer pixel accuracy. Hence, for one-level DWT, the MVs in (11) are reduced to a half scale. The MV error can be expressed as

$$\Delta d3 = (\widetilde{d}_x, \widetilde{d}_y) - \left(\frac{d_x'}{2}, \frac{d_y'}{2}\right), \tag{17}$$

where $(\widetilde{d}_x, \widetilde{d}_y)$ is the true MVs of low-pass subband. So the PSD of the low-pass subband prediction error can be derived as

$$\Phi_{\widehat{e}_l\widehat{e}_l}(\omega) = 2\Phi_{S_lS_l}(\omega)\left[1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d3}}\right] + \widetilde{\theta}, \quad \omega \in \omega_l. \tag{18}$$

According to (13)–(18), the rate difference between the DPCM coding of the spatial auxiliary information and the low-pass subband of the full-resolution frame which is encoded by MCE-based WZVC can be derived as

$$\begin{aligned}&\Delta R_l(\theta) \\ &= R_l(\widetilde{\theta}) - R_l(\widehat{\theta}) \\ &= \frac{1}{8\pi^2}\int_{\omega_x\in\omega_l}\int_{\omega_y\in\omega_l} \log_2\frac{\Phi_{\widetilde{e}_l\widetilde{e}_l}(\omega)}{\widetilde{\theta}}d\omega \\ &\quad - \frac{1}{8\pi^2}\int_{\omega_x\in\omega_l}\int_{\omega_y\in\omega_l} \log_2\frac{\Phi_{\widehat{e}_l\widehat{e}_l}(\omega)}{\widehat{\theta}}d\omega \\ &= \frac{1}{8\pi^2}\int_{\omega_x\in\omega_l}\int_{\omega_y\in\omega_l} \log_2\frac{\widehat{\theta}}{\widetilde{\theta}}\frac{\Phi_{\widetilde{e}_l\widetilde{e}_l}(\omega)}{\Phi_{\widehat{e}_l\widehat{e}_l}(\omega)}d\omega \\ &= \frac{1}{8\pi^2}\int_{\omega_x\in\omega_l}\int_{\omega_y\in\omega_l} \log_2\frac{\widehat{\theta}}{\widetilde{\theta}}\frac{\left(1-e^{-(1/2)\omega^T\omega\sigma^2_{\Delta\text{MV1}}}\right)+\widetilde{\theta}/2\Phi_{S_lS_l}}{\left(1-e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d3}}\right)+\widehat{\theta}/2\Phi_{S_lS_l}}d\omega.\end{aligned} \tag{19}$$

Since it is assumed that the PSD of spatial signal is much larger than the PSD of quantization noise signal, the function (19) can be simplified as

$$\Delta R_l(\theta) = \frac{1}{8\pi^2}\int_{\omega_x\in\omega_l}\int_{\omega_y\in\omega_l} \log_2\frac{\widehat{\theta}}{\widetilde{\theta}}\frac{\left(1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta\text{MV1}}}\right)}{\left(1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d3}}\right)}d\omega. \tag{20}$$

According to (10) and (20), it can be derived that the overall rate saving ΔR is

$$\begin{aligned}\Delta R(\theta) &= \Delta R_l(\theta) + \Delta R_f(\theta) \\ &= \frac{1}{8\pi^2}\int_{\omega_x\in\omega_l}\int_{\omega_y\in\omega_l} \log_2\frac{\widehat{\theta}}{\widetilde{\theta}}\frac{\left(1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta\text{MV1}}}\right)}{\left(1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d3}}\right)}d\omega \\ &\quad + \frac{1}{8\pi^2}\int_{\omega_x}\int_{\omega_y} \log_2\left(\frac{1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d2}}}{1 - e^{-(1/2)\omega^T\omega\sigma^2_{\Delta d1}}}\right)d\omega.\end{aligned} \tag{21}$$

The first part of (21) can be considered as the overhead by the auxiliary information coding. The second part of (21) is the coding gain from the spatial auxiliary information-aided motion-compensated extrapolation.

3.2.2. Numerical Results. The rate saving for SA-WZVC versus MCE-based WZVC is examined as follows. According to the statistics of displacement error and the quantization noise's PSD ratio, the rate difference is obtained by (21). The numerical results of theoretical analysis are shown in Table 1 where different qualities of auxiliary information are used in SA-WZVC. This results in different displacement errors and different overheads consumed by the auxiliary information coding. Therefore, different rate savings can be achieved.

In the simulation, *Foreman* CIF sequence is used and twenty WZ frames are encoded. One-level 9/7 wavelet decomposition is adopted to generate spatial auxiliary information. The quality of key frames in SA-WZVC and MCE-based WZVC is the same. When the quantization scheme of MCE-based WZVC is determined, the quantization error $\hat{\theta}$ introduced into low-pass subband is confirmed. The PSD ratio $\hat{\theta}/\tilde{\theta}$ of quantization error in (21) is only determined by the quantization error of auxiliary information $\tilde{\theta}$. SNR represents the correlation of MVs generated by MCE method and MVs generated by SA-MCE method. It is calculated as follows:

$$\mathrm{SNR} = 10\log_{10}\frac{\sigma^2_{\Delta d1}}{\sigma^2_{\Delta d2}}. \tag{22}$$

For the same reason, when the quantization error of WZ frame and the quantization error of key frame are determined, MVs generated by MCE method is constant too. So the SNR is only affected by the variance of MVs generated by SA-MCE. Also, ΔR_f is the rate difference between the spatial-aided WZ coding and MCE-based WZ coding which is defined in (10); ΔR is the overall rate saving defined as (21) that comprises the overhead coding and the rate saving of WZ coding. From the simulation result it can be derived that there exists a tradeoff between the auxiliary information coding and the WZ frame coding. As the quantization error of auxiliary information coding decreased, the SNR increases and the rate saving of WZ coding ΔR_f increases. This phenomenon illustrates that if more bits are allocated to auxiliary information coding, the accuracy of MVs generated with the help of the high-quality auxiliary information is improved. The variance of MV error $\sigma^2_{\Delta d2}$ decreases. Therefore, the rate saving of WZ coding ΔR_f is increased. However, the overhead brought by the auxiliary information coding is also increased. The overall rate saving is decreased.

On the contrary, if the quality of auxiliary information decreases, both the accuracy of MVs and the rate saving of WZ coding ΔR_f are decreased. The quality of auxiliary information is important that it can affect the coding tradeoff. It can be concluded that if the strategy of bit allocation is optimum, a promising coding gain can be achieved.

4. Experimental Results and Analysis

In this section, the proposed scheme is implemented to verify the coding efficiency of the spatial-aided low-delay WZ coding. The key frames are H.264/AVC-intra-coded using the reference software JM 9. The spatial auxiliary information is generated by applying DWT decomposition to the original frames and the DWT is implemented with biorthogonal 9/7 filter. The entropy coding method adopted in DPCM coding of spatial auxiliary information is CA-VLC in JM 9. For the low-delay WZ coding of the whole frame, as described in Section 2.3, DCT domain WZ coding scheme is used. A rate-compatible punctured turbo encoder (RCPT) is adopted as Slepian-Wolf codec and the acceptable bit-error rate at the decoder is set to 10^{-3}. The parameter of Laplacian distribution model is obtained by offline fitting the difference between the original frame and its side information frame. Due to different distributions, the parameters of each bit-plane may have different values. For various sequences, different parameters of Laplacian distribution model are also obtained by offline training.

Foreman, News, and Tempete sequences at CIF resolution are used in testing. In each sequence, 168 frames are encoded and the coding structure is I-W-,$\cdots$-,W-I. The QP for DPCM coding of spatial auxiliary information is equals to the QP of key frames minus two. Five different QPs are chosen for key frame coding: 20, 24, 28, 32, and 36.

4.1. Evaluation of Spatial-Aided Wyner-Ziv Video Coding. The overall RD performance of the "SA-WZVC" is compared with that of a scheme proposed in [7]. In Figures 3(a), 3(b), and 3(c), "SA-WZVC" denotes the proposed spatial-aided WZ coding. One level 2D-DWT with biorthogonal 9/7 filter is applied to generate auxiliary information. The GOP size adopted in the simulation is 6. The scheme proposed in [7] is implemented and it is denoted as "Hybrid Intra/WZVC" in Figures 3(a), 3(b), and 3(c). The auxiliary generation method of the "Hybrid Intra/WZVC" is in spatial domain. Compared with the RD performance of "Hybrid Intra/WZVC," our method also achieves a promising improvement. The quality of the key frames used in our proposed methods and in Hybrid Intra/WZVC scheme proposed in [7] remain the same. The curve of "H.264 Intra" indicates the results of H.264/AVC intra-frame coding. Compared with the overall RD performance of the intra-frames coding and DPCM coding, it can be observed that the proposed method efficiently improves the rate distortion performance of WZVC in low-delay application.

The ratio of the bit-rate used in key frame coding, auxiliary information coding and WZ coding are presented in Tables 2(a), 2(b), and 2(c), respectively.

In Table 2, QP_k represents the quantization parameter of key frame coding. The QP for DPCM coding of spatial auxiliary information is equals to the QP of key frames minus two. According to Tables 2(a), 2(b), and 2(c), at the high bit-rate point, most percent of bit-rate is consumed by intra-coding of key frames and the auxiliary information coding. The WZ frame coding takes a much low percent. At the low bit-rate point, the rate consumed by WZ coding cannot be

Table 1: Numerical Results: Foreman@CIF.

Key QP	$\hat{\theta}/\tilde{\theta}$	$\sigma^2_{\Delta d1}$	$\sigma^2_{\Delta d2}$	SNR	ΔR_f	ΔR
20	1.0494	3.3521	1.9967	2.25	−0.0329	0.1083
	0.8815	3.3521	2.0394	2.16	−0.0317	−0.0130
	0.7557	3.3521	2.1237	1.98	−0.0241	−0.1133
	0.6385	3.3521	2.1749	1.88	−0.0193	−0.2272
	0.5521	3.3521	2.2628	1.71	−0.0118	−0.3229
24	0.8820	2.6395	1.7153	1.87	−0.0292	0.0324
	0.7423	2.6395	1.7960	1.67	−0.0196	−0.0787
	0.6481	2.6395	1.8487	1.55	−0.0134	−0.1680
	0.5758	2.6395	1.9222	1.38	−0.0105	−0.2492
	0.5000	2.6395	2.0777	1.04	0.0033	−0.3354

Table 2: The ratio of the bit-rate in SA-WZVC, GOP = 6.

(a)

Foreman@CIF GOP size = 6	Over all PSNR	Over all bit-rate (kbps)	Bit-rate ratio of key frame coding (%)	Bit-rate ratio of auxiliary information coding (%)	Bit-rate ratio of WZ coding (%)
$QP_k = 20$	36.6	1796.47	38.25	60.78	0.97
$QP_k = 24$	35.28	1176.42	39.34	59.01	1.65
$QP_k = 28$	33.62	774.04	39.76	56.64	3.6
$QP_k = 32$	31.8	505.92	38.89	52.5	8.6
$QP_k = 36$	29.87	348.74	36.57	42.69	20.74

(b)

News@CIF GOP size = 6	Over all PSNR	Over all bit-rate (kbps)	Bit-rate ratio of key frame coding (%)	Bit-rate ratio of auxiliary information coding (%)	Bit-rate ratio of WZ coding (%)
$QP_k = 20$	39.53	861.02	66.53	31.95	1.51
$QP_k = 24$	37.79	631.49	66.98	30.07	2.95
$QP_k = 28$	35.71	459.05	67.66	26.89	5.45
$QP_k = 32$	33.53	337.22	64.84	23.54	11.62
$QP_k = 36$	31.46	255.4	59.73	19.03	21.25

(c)

Tempete@CIF GOP size = 6	Over all PSNR	Over all bit-rate (kbps)	Bit-rate ratio of key frame coding (%)	Bit-rate ratio of auxiliary information coding (%)	Bit-rate ratio of WZ coding (%)
$QP_k = 20$	32.15	2541.52	48.09	49.91	1.99
$QP_k = 24$	31.08	1934.92	48.45	48.5	3.05
$QP_k = 28$	29.68	1398.43	49.76	45.02	5.22
$QP_k = 32$	27.96	964.04	49.58	40.01	10.41
$QP_k = 36$	26.22	673.7	46.48	31.07	22.45

ignored. Hence, the gain in $R(D)$ performance comes from a combination of WZ coding and the spatial-aided motion extrapolation.

4.2. Evaluation the Performance for Varying GOP Size. Considering the low-delay application scenarios, the simulation of SA-WZVC adopting longer GOP size is performed. In Figures 4(a), 4(b), and 4(c), the case of GOP size 12 is compared with GOP size 6. The test sequence and the quantization parameters of key frame coding are the same with the former simulations. One-level 2D-DWT with biorthogonal 9/7 filter is applied to generate auxiliary information.

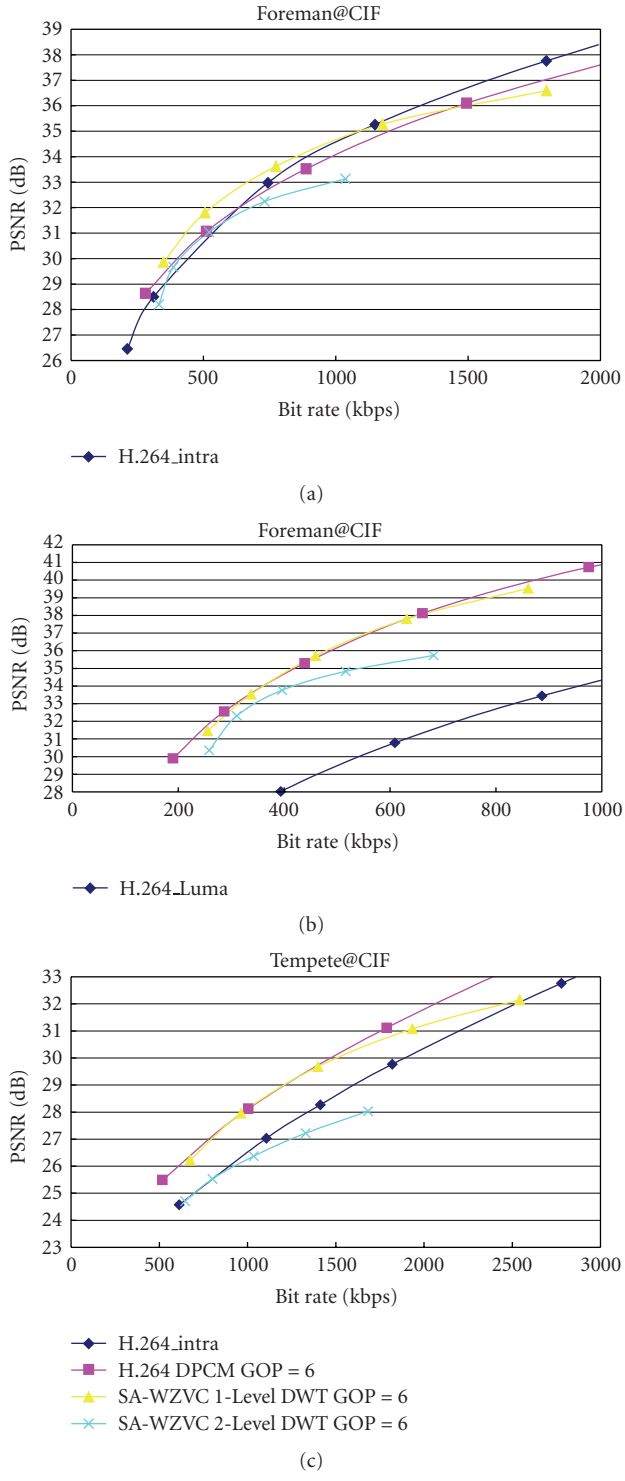

Figure 6: Overall RD performance comparisons for multilevel DWT.

From the simulation results, it can be concluded that longer GOP size degrades the RD performance for the test sequence with high motion such as *Foreman* and *Tempete*. In fact, the quality of key frames is very important for the overall RD performance (including both key frames and WZ frames). To investigate this phenomenon, the frame by frame PSNR distribution of the decoded frames and the distribution of bit-rate in *Foreman* sequence are presented in Figures 5(a) and 5(b). According to Figure 5(a), it is found that the quality of WZ frame located in forward position is better than the quality of WZ frame located in backward position in one GOP. However, the bits consumed by WZ coding of backward frames increase compared to the bit-rate of forward WZ frame according to Figure 5(b). It is because that as the frame number increases in one GOP, the quality of reference frames decreases, and this results in the degradation of the SI quality. To recover more errors between SI and the original signals, it has to cost more bits in WZ decoding. Therefore, the performance of WZVC in long GOP size case might decrease. Key frame has to be refreshed in a proper period. For the sequences with smooth motion, such as *News*, longer GOP size can bring improvement in RD performance. How to find a proper GOP size for low-delay WZVC is our future research topic.

4.3. Experiments with Multilevel DWT. If more than one level wavelet decomposition is carried out, the auxiliary information with smaller resolution is generated and it can produce negligible overhead from the auxiliary information for the whole system. The simulation of SA-WZVC using two-level decomposition has been done. In this case, the lowest-pass subband with the resolution of 88×72 is transmitted as auxiliary information. The higher-resolution SI is extrapolated with the aid of lower subband by using the SA-MCE method. The higher resolution frames are successfully refined by WZ coding methods. The RD performance is shown in Figures 6(a), 6(b), and 6(c). Comparing with one-level DWT decomposition, there is a performance loss in two-level DWT. By a carefully study, it is found that the correlation between the SI and the original information becomes more weaker. This phenomenon attributes to two factors: the energy contained in auxiliary information decreases due to the multilevel DWT and the accuracy of motion information is diminished since the MVs are generated with the aid of the imperfect auxiliary information. The correlation decreasing induces the increasing of rate cost in WZ coding. This cost cannot compensate the rate reduction in overhead coding.

5. Conclusions

In this paper, a spatial-aided low-delay WZ coding scheme has been presented. In this scheme, the low-pass subband of WZ frame generated by DWT is used as the spatial auxiliary information and encoded by DPCM. At the decoder, the spatial auxiliary information is decoded first. By performing motion estimation on the upsampled spatial auxiliary information, more accurate MVs are obtained comparing with MCE-based SI generation. This improvement enables us to implement a high-efficiency low-delay WZ coding. In our further study, a more general analysis will be considered at the full scale only. The low-pass subband is coded and transmitted as auxiliary information. The high-pass subband could be encoded independently by spatial-aided low-delay WZVC method. In this case, all of the impacts brought

by decimation, subsequent interpolation, and simple-coarse quantization could be considered at full scale in a more general manner. Moreover, to fully explore the characteristic of the proposed SA-WZVC in low-delay applications, the case of longer GOP size and the case of one I-frame followed by all WZ frames will be studied and realized in further research. How to find a proper GOP size for low-delay WZVC is also a future research topic.

Acknowledgments

This work was supported in part by the National Science Foundation of China (60736043 and 60672088) and Major State Basic Research Development Program of China (973 Program, 2009CB320905).

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[3] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in *Visual Communications and Image Processing*, vol. 5308 of *Proceedings of SPIE*, pp. 520–528, San Jose, Calif, USA, January 2004.

[4] A. Aaron and B. Girod, "Wyner-Ziv video coding with low encoder complexity," in *Proceedings of the Picture Coding Symposium (PCS '04)*, San Francisco, Calif, USA, December 2004.

[5] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, vol. 5, pp. 3097–3100, Singapore, October 2004.

[6] S. Rane, A. Aaron, and B. Girod, "Systematic lossy forward error protection for error-resilient digital video broadcasting—a Wyner-Ziv coding approach," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, vol. 5, pp. 3101–3104, Singapore, October 2004.

[7] D. Agrafiotis, P. Ferré, and D. R. Bull, "Hybrid key/Wyner-Ziv frames with flexible macroblock ordering for improved low delay distributed video coding," in *Visual Communications and Image Processing*, vol. 6508 of *Proceedings of SPIE*, pp. 1–7, San Jose, Calif, USA, January-February 2007.

[8] E. Martinian, A. Vetro, J. S. Yedidia, J. Ascenso, A. Khisti, and D. Malioutov, "Hybrid distributed video coding using SCA codes," in *Proceedings of the 8th IEEE Workshop on Multimedia Signal Processing (WMSP '06)*, pp. 258–261, Victoria, Canada, October 2006.

[9] N. Mehrseresht and D. Taubman, "A flexible structure for fully scalable motion-compensated 3-D DWT with emphasis on the impact of spatial scalability," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 740–753, 2006.

[10] R. Xiong, J. Xu, F. Wu, S. Li, and Y.-Q. Zhang, "Subband coupling aware rate allocation for spatial scalability in 3-D wavelet video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 10, pp. 1311–1324, 2007.

[11] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Wyner-Ziv video coding based on set partitioning in hierarchical tree," in *Proceedings of the International Conference on Image Processing (ICIP '06)*, pp. 601–604, Atlanta, Ga, USA, October 2006.

[12] C. Tang, N.-M. Cheung, A. Ortega, and C. S. Raghavendra, "Efficient inter-band prediction and wavelet based compression for hyperspectral imagery: a distributed source coding approach," in *Proceedings of the Data Compression Conference (DCC '05)*, pp. 437–446, Snowbird, Utah, USA, March 2005.

[13] J. E. Fowler, M. Tagliasacchi, and B. Pesquet-Popescu, "Wavelet-based distributed source coding of video," in *Proceedings of the 13th European Signal Processing Conference*, Antalya, Turkey, September 2005.

[14] M. Grangetto, E. Magli, and G. Olmo, "Context-based distributed wavelet video coding," in *Proceedings of the 7th IEEE Workshop on Multimedia Signal Processing (WMSP '05)*, pp. 1–4, Shanghai, China, October-November 2005.

[15] M. Wu, A. Vetro, J. Yedidia, H. Sun, and C. W. Chen, "A study of encoding and decoding techniques for syndrome-based video coding," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '05)*, vol. 4, pp. 3527–3530, Kobe, Japan, May 2005.

[16] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 7, pp. 1140–1154, 1987.

[17] B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 604–612, 1993.

[18] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70–84, 1999.

[19] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 173–183, 2000.

[20] M. Flierl, T. Wiegand, and B. Girod, "Rate-constrained multihypothesis prediction for motion-compensated video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 11, pp. 957–969, 2002.

[21] Z. Li, L. Liu, and E. J. Delp, "Rate distortion analysis of motion side estimation in Wyner-Ziv video coding," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 98–113, 2007.

[22] M. Tagliasacchi and S. Tubaro, "Hash-based motion modeling in Wyner-Ziv video coding," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '07)*, vol. 1, pp. 509–512, Honolulu, Hawaii, USA, April 2007.