Progressive Side Information Refinement Algorithm for Wyner-Ziv Codec

Chan-Hee HAN†, Si-Woong LEE†, Hamid GHOAMHOSSEINI††, Nonmembers, and Yun-Ho KO†††(a), Member

SUMMARY In this paper, side information refinement methods for Wyner-Ziv video codec are presented. In the proposed method, each block of a Wyner-Ziv frame is separated into a predefined number of groups, and these groups are interleaved to be coded. The side information for the first group is generated by the motion compensated temporal interpolation using adjacent key frames only. Then, the side information for remaining groups is gradually refined using the knowledge of the already decoded signal of the current Wyner-Ziv frame. Based on this basic concept, two progressive side information refinement methods are proposed. One is the band-wise side information refinement (BW-SIR) method which is based on transform domain interleaving, while the other is the field-wise side information refinement (FW-SIR) method which is based on pixel domain interleaving. Simulation results show that the proposed methods improve the quality of the side information and rate-distortion performance compared to the conventional side information refinement methods.

key words: Wyner-Ziv video codec, distributed video coding, side information refinement, band-wise side information refinement, field-wise side information refinement

1. Introduction

Current video coding solutions, such as MPEG or ITU-T H.26x standards, perform well for such applications as broadcasting and streaming services, where a video is encoded once and decoded several times [1]. In this coding framework, the encoder architecture is based on the combination of motion estimation tools with DCT transform, quantization and entropy coding in order to exploit the temporal, spatial and statistical redundancies in a video sequence. In this framework, the encoder has a higher computational complexity than the decoder, typically 5 to 10 times more complex [2]. However, this architecture is being challenged by several emerging applications such as wireless video surveillance, multimedia sensor networks, wireless PC cameras and mobile camera phones. These applications have different requirements from those targeted by traditional video delivery systems. For example, in wireless video surveillance systems, low-cost encoders are required since there are a huge number of encoders and only one or few decoders [3]. For this reason, distributed video coding (DVC) or Wyner-Ziv video coding which allows very low complex encoding is well-suited for these applications. According to the Slepian-Wolf theorem [4], it is possible to achieve the same rate as a joint encoding system by independent encoding but joint decoding of two statistically dependent signals. The Wyner-Ziv theorem [5] extends the Slepian-Wolf theorem to a lossy case, which becomes the key theoretical basis of DVC [6].

In Wyner-Ziv (WZ) video codec, some frames of the video sequence are intra-coded (called key frames), while the others (called Wyner-Ziv frames, or WZ frames) are coded using the WZ encoding and decoding scheme. At the decoder, intra frames are interpolated to generate an estimation of the WZ frames, which is then combined with the information sent by the encoder to reconstruct the WZ frames. From the decoder view point, the information sent by the encoder is known as main signal, and the estimation generated at the decoder is known as side information (SI).

The quality of SI has a significant influence on the RD (rate and distortion) performance of WZ codec. The decoder having more accurate SI requests fewer parity bits from the encoder, so the bitrates can be reduced to achieve a target quality. Thus, how to generate more accurate estimate of a WZ frame, i.e., the SI (side information) at the decoder is one of the most important issues of WZ codec. Recently, many efforts have been contributed to improve the performance of the SI generation (SIG). These methods can be classified into two main groups, namely, the MCTI (motion compensated temporal interpolation) and the HME (hash-based motion estimation) approaches.

Figure 1 shows the architecture of the MCTI-based WZ video codec. The MCTI methods [7]–[13] generate SI by compensating with the motion between past and future key frames. However, MCTI methods are based on the conventional methods such as the frame rate up-conversion [14], [15], so the quality of SI is limited to the performance of existing research. Especially, the MCTI methods show poor performance in the WZ frames including occluded region or non-uniform motion between the WZ frame and both key frames, since in these cases the generated SI using only motion between two key frames cannot be a correct estimation of the WZ frame.

Figure 2 shows the architecture of the HME-based WZ video codec. In the HME-based approach [16], [17], the encoder sends to the decoder some extra information about the current frame, in order to aid the decoder to generate more correct side information by the motion estimation process. Therefore, HME methods increase the complexity and bi-
Fig. 1 MCTI-based WZ video codec.

Fig. 2 HME-based WZ video codec.

Fig. 3 General architecture of PDWZ codec with SIR.

Fig. 4 General architecture of TDWZ codec with SIR.


trates of the WZ encoder because of the additional hash bits. However, this is not desirable in the sense that the structure of the encoder in WZ video coding has to be very simple. So most of the research for WZ video codec is focused on the MCTI-based scheme.

More recently, side information refinement (SIR) methods for MCTI-based WZ video codec are proposed in order to get more accurate SI without requiring extra information from the encoder. Figure 3 and Fig. 4 show general pixel domain and transform domain WZ codec architecture including SIR procedure, respectively. The SIR procedure refines the SI iteratively after each partial decoding of the predefined basic unit of the current WZ frame. And the refined SI, the more accurate estimate is used for the remaining decoding of the current WZ frame to improve the rate-distortion performance of WZ codec. Commonly, in case of the pixel domain WZ (PDWZ) codec, the basic unit is each bitplane of the quantized pixel values for a WZ frame. And, in case of the transform domain WZ (TDWZ) codec, the basic unit is each bitplane of a band which is a bunch of collocated transform coefficients.

As shown in Fig. 3, previous pixel domain WZ (PDWZ) codec with SIR continuously improves the side information as close as possible to the current WZ frame after each partial decoding of the current frame. In [18], pre-reconstructed version of the current WZ frame is used to enhance SI for the remaining bitplanes to be decoded. This approach uses the pre-reconstructed image and the adjacent key frames in order to refine the motion vectors and thus to obtain a new and improved version of the decoded and SI frames. In [19], the SI is refined using previously decoded bitplanes such as [18]. But the SIR process of this paper uses only past key frame to predict the current WZ frame in order to develop a low-delay WZ codec. In [20], a spatial-temporal refinement algorithm extending [19] is proposed to iteratively improve the initial SI obtained by motion extrapolation.

Although most research related to SIR has been implemented for PDWZ codec, it is well known that transform based architectures which can consider the spatial redundancy are more efficient. Recently, some TDWZ-based SIR methods have been presented. In [21], DC band is decoded with initial SI and then DC domain motion estimation is performed between a DC frame obtained by the decoded DC band and the oversampled DC (OS DC) frames consisting of DC values of all possible transform blocks of the key frames. Finally, initial SI is refined using the motion vectors obtained by the DC domain motion estimation and remaining AC bands is decoded with the refined SI. However the motion vectors obtained by the motion estimation using only DC values are unreliable and thus the improvement of rate-distortion performance is limited. [22] and [23] are most recent TDWZ-based SIR methods that are implemented with state-of-the-art TDWZ codec technologies.

In [22], the SI for the next band is refined using already decoded bands after decoding of each band. A block of the previous SI is refined by weighted averaging of the candidate blocks of the previous SI. The candidate blocks are selected based on the error matching between blocks of the
previous SI within a search range and a pre-reconstructed pixel block as shown in Fig. 5. This approach is efficient from the viewpoint of the trade-off between the performance and the complexity since the motion estimation using key frames is not performed. The [23] extending [22] proposed a statistical motion learning to improve the correlation noise distribution modeling as well as the SIR. However, despite good trade-off between the codec performance and the complexity, SIR methods of [22] and [23] have obvious limitation in an image sequence where there is occluded region or non-uniform motion because the iterative refinement procedure is progressed using the information of the previous SI only without the information of key frames. In other words, their refinements are too much dependent on the initial SI so they cannot show good performance when the initial SI is a bad estimate caused by occluded region or non-uniform motion.

In this paper we propose two SIR methods to improve the RD performance of the overall WZ video codec. The first one is band-wise SIR (BW-SIR) method that is applicable to the TDWZ codec and the other is field-wise SIR (FW-SIR) method that is applicable to the PDWZ codec. In the BW-SIR, each $4 \times 4$ transform block of a WZ frame is separated into 16 groups and these groups are interleaved to be encoded in a similar way as [22] and [23]. Compared with these conventional TDWZ-based SIR methods, the main contribution and difference of the proposed method is to make use of adjacent key frames at each refinement procedure as shown in Fig. 6. This makes it possible to achieve more accurate refinement especially when there is occluded region or non-uniform motion. The idea of using key frames in SIR is rejected in [22] due to its computational complexity. However, increase in the decoder complexity is not important in the distributed video coding (DVC). The basic concept and architecture of the DVC is based on low complexity encoder and high complexity decoder. We propose an efficient refinement scheme to increase the rate-distortion performance while preventing from increasing in computational complexity.

In the FW-SIR, a WZ frame is separated into 2 groups of top and bottom fields and a bottom field is coded after the coding of the relevant top field. The decoding of top field of a WZ frame is performed using initial SI generated by adjacent key frames only, and then the SI for the bottom field is refined using motion estimation and compensation tool between previously decoded top field information and adjacent key frames. While the conventional PDWZ-based SIR methods use a bitplane-level approach as illustrated in Fig. 3, the proposed method use a field-level interleaving approach. In addition, we propose a refinement framework that combines the conventional bitplane-level approach and the FW-SIR in order to achieve better RD performance. This paper is organized as follows. Section 2 and 3 presents the BW-SIR and FW-SIR algorithm, respectively. Section 4 presents the test conditions, the performance evaluation results, and their analysis of each method. Finally we conclude the paper in Sect. 5.

2. Band-Wise Side Information Refinement

2.1 Codec Architecture Using BW-SIR

Figure 7 shows the architecture of TDWZ codec using the proposed BW-SIR. A WZ frame is transformed using the $4 \times 4$ discrete cosine transform (DCT) whose coefficients are composed of 16 frequency bands. All coefficients with the same frequency over a WZ frame are grouped together to form a specific band. The frequency bands are successively coded. In other words, after the DC band including all DC coefficients of the WZ frame is coded, and then the AC bands are coded according to the zigzag scanning order. Thus, the block data is separated into 16 groups and intrinsically interleaved in the TDWZ codec as shown in Fig. 7.

SI for DC band is generated in the same way as the conventional SIG (Side Information Generation) method [7], since the information for the current WZ frame is not available at this point in the decoder. It is based on the MCTI using the past and future key frames. In order to achieve higher precision in the predicted signal, the MCTI method requires that the basic assumption of the smooth and uniform motion
between key frames has to be satisfied. However, this assumption has two failure modes: occluded region and non-uniform motion. Thus, in these cases, the initial SI contains high prediction error which leads to a severe degradation in overall RD efficiency.

To alleviate the problem, the initial SI is used for only DC band reconstruction procedure in the proposed method. For the reconstruction of remaining AC bands, SI is progressively refined using the previously decoded frequency bands of the current WZ frame. As the decoded bands are accumulated, the quality of SI continues to improve and less parity bits are requested to the encoder. Also it is important to note that the proposed BW-SIR does not require the increase of the complexity of the encoder.

2.2 BW-SIR Process

Figure 8 shows the proposed BW-SIR process. When each frequency coefficient is reconstructed in the WZ decoder using the received parity bits, it is compared with the corresponding frequency coefficient in SI to verify the validity of each block in SI. Those blocks having large difference between them are classified as error blocks. Then, for such blocks, the coefficient in SI is replaced with the reconstructed one, and a motion compensated side information refinement (MC-SIR) process is applied to obtain a new SI with higher quality. This SIR process is iteratively applied to all the remaining AC bands except for the first DC band.

2.2.1 Error Block Detection (EBD) & Error Block Correction (EBC)

As previously described, the error blocks in SI are detected based on the absolute difference between the decoded coefficient and the corresponding one in SI as follows:

$$|\text{Coef}_{k}^{\text{decoded}} - \text{Coef}_{k}^{\text{SI}}| > \varepsilon_k$$

where $\text{Coef}_{k}^{\text{decoded}}$ and $\text{Coef}_{k}^{\text{SI}}$ are the $k$-th ($k = 0, \ldots, 15$) zig-zag ordered decoded coefficient and the $k$-th SI coefficient in a block of a WZ frame, respectively. $\varepsilon_k$ is a predetermined threshold. Since $\text{Coef}_{k}^{\text{SI}}$ is an estimate of $\text{Coef}_{k}^{\text{decoded}}$, the large difference between them means that the block is erroneously predicted and thus more accurate estimation is needed.

The threshold $\varepsilon_k$ in Eq. (1) plays the role of preventing the excessive increment of computational burden by restricting SIR process to error blocks only. We can control the required amount of computations by adjusting its value. As the $k$ increases, generally the magnitude of the $k$-th coefficient decreases. So it is more desirable in the view point of computational efficiency to use smaller $\varepsilon_k$ in the higher band. In our experiment, it is decided statistically by averaging the difference values over $N$ test sample blocks as follows:

$$\varepsilon_k = \alpha \frac{1}{N} \sum_{n=0}^{N-1} |\text{Coef}_{k}^{\text{decoded}}[n] - \text{Coef}_{k}^{\text{SI}}[n]|$$

where $|\text{Coef}_{k}^{\text{decoded}}[n] - \text{Coef}_{k}^{\text{SI}}[n]|$ is the absolute difference of the $n$-th test sample block and $\alpha$ is a parameter to control the amount of computations.

When a block of the previously generated SI is determined as an error block after the $k$-th coefficient decoding, it is corrected by replacing lower band transform coefficients up to $k$-th coefficient with the decoded ones in the EBC (Error Block Correction) procedure. These updated coefficients enhance the motion estimation performance in the following motion compensated SIR procedure.

2.2.2 Motion Compensated Side Information Refinement (MC-SIR)

Once the error blocks are corrected with the received coefficients, additional ME processes are performed as shown in Fig. 6. Note that, contrary to the initial SI generation where only motion estimation between adjacent key frames is performed, direct motion estimation between the current pre-reconstructed WZ frame and key frames is available in SIR process. Therefore, it is proposed in this paper to use 3 modes of motion estimation for each $4 \times 4$ error block as
follows:

$$\min_{MV^T \in S} MAD\left(\hat{B}, K_P\left(MV^F\right)\right)$$ (3)

$$\min_{MV^B \in S} MAD\left(\hat{B}, K_N\left(MV^B\right)\right)$$ (4)

$$\min_{MV^B \in S} MAD\left(\hat{B}, \frac{K_P\left(MV^B\right) + K_N\left(MV^B\right)}{2}\right)$$ (5)

where \(MV^F\), \(MV^B\) and \(MV^B_i\) are forward, backward and bidirectional motion vectors, respectively. \(S\) is the search range. \(K_P\) and \(K_N\) are the past and future key frames respectively. \(\hat{B}\) is the block corrected at the previous stage.

Comparing the MADs of three motion vectors, the optimal motion vector is selected and used to generate the refined SI. Uni-directional motion compensation such as forward and backward motion compensations in Eq. (3) and Eq. (4) is especially efficient in predicting the blocks in occluded regions. The bi-directional motion compensation in Eq. (5) is efficient in the sequences where linear illumination changes such as fade-in or fade-out occur. The received coefficients contained in \(\hat{B}\) play the role of selecting the best compensation mode for each block, which results in enhanced SI for remaining bands.

3. Field-Wise Side Information Refinement

3.1 Codec Architecture Using FW-SIR

Figure 9 shows the architecture of the PDWZ video codec using the proposed FW-SIR. At the encoder, a WZ frame is separated into top and bottom fields as the field coding scheme of the conventional video codec, and each field is interleaved to be coded. The top field is coded first, and then the bottom field is coded.

At the decoder, SI for the top field is generated in the same way as the conventional SIG method [7]. It is based on the MCTI using the past and future key frames. In the SI generation process for the bottom field, the top field signal of the current WZ frame is available, since it is already reconstructed. Therefore, accurate motion vectors for the bottom field can be computed by matching the decoded top field signal with the key frames. This is the main reason why the proposed field-wise SIR method can improve the quality of the side information for the bottom field.

While the conventional PDWZ-based SIR methods utilize a bitplane-level interleaving only for refinement procedure as shown in Fig. 3, the proposed method may use a field-level interleaving additionally as shown in Fig. 9. Therefore two refinement schemes are possible as illustrated in Fig. 9. The first refinement scheme makes use of field-level interleaving only at the decoder as shown in Fig. 9 (a). In this scheme the SI for the bottom field is refined using motion estimation and compensation tool between previously decoded top field information and adjacent key frames. The second refinement scheme utilizes both field-level interleaving and bitplane-level interleaving simultaneously. In the second scheme the top field of the first bitplane is decoded using the initial SI generated by adjacent key frames. Then the SI for the bottom field of the first bitplane is refined using the proposed FW-SIR method. After decoding the first biplane, the SI for the remaining bitplane is refined continuously using the conventional PDWZ-based SIR method such as [18].

3.2 FW-SIR Process

Figure 10 shows the proposed SIR process for bottom fields. It is very similar to BW-SIR process in Sect. 2.2 excepting that decoded top field data instead of decoded transform coefficients are used for error block detection and the iteration procedure is not needed. When a top field is reconstructed in the proposed WZ decoder using the received parity bits, it is compared with the top field in SI on a per block basis to verify the validity of each block in SI. Those blocks having larger differences than a threshold \(\varepsilon\) are classified as error blocks as follows:

$$SAD_i = \sum_{(x,y) \in B_i} |I_{decoded}(x,y) - I^{SI}(x,y)| > \varepsilon$$ (6)

where \(I_{decoded}(x,y)\) and \(I^{SI}(x,y)\) are pixel values of decoded signal and SI at coordinates \((x,y)\), respectively. \(B_i\) represents \(i\)-th block in the top field. The threshold \(\varepsilon\) is determined as follows:

$$\varepsilon = \alpha \frac{1}{M} \sum_{i=0}^{M-1} SAD_i$$ (7)
where $M$ is the number of all blocks in the top field and $\alpha$ is a parameter to control the amount of error blocks. Then, the decided error blocks are corrected by replacing their top field pixels with the reconstructed ones.

Once the error blocks are corrected with the received top field signal, a new motion compensated side information refinement (MC-SIR) process is applied according to Eq. (3)–Eq. (5). In the ME process, the decoded top field signal drives the accurate motion vectors for the current block to be estimated so that more accurate SI can be generated for the bottom field.

### 4. Simulation Results

In order to evaluate the performance of the proposed methods, the WZ video codecs based on the structure in Fig. 7 (for BW-SIR) and Fig. 9 (for FW-SIR) are implemented in our experiments. For the quantization module, the well known uniform quantizer of $2^M$ levels where $M$ is determined depending on the target quality for the WZ frame is used. The quantization indices are then organized in $M$ bit planes and each bit plane is fed to the channel encoder. For the channel coder, the rate-compatible LDPC Accumulate (LDPCA) code [24] is used, and for the virtual channel model in the WZ decoder, a laplacian distribution model based on [25] is used to model the correlation noise. In addition, a reconstruction model based on [26] is used for the reconstruction module.

The detailed test condition and the conventional methods used for the RD performance comparison are summarized as follows.

- **Quantization and RD points**
  - BW-SIR : Eight RD points. Figure 11 shows the eight $4 \times 4$ quantization matrices. The value in the matrix denotes the number of bits for representing the reconstruction levels of each DCT band.
  - FW-SIR : Five RD points. A Quantizer having $2^M$ reconstruction levels is applied for all pixel values of a frame, where $M$ for $Q_i (i = 1, 2, 3, 4, 5)$ is $i$.
- **LDPCA code length**
  - BW-SIR : 1584. It is determined based on the fact that the number of coefficients in each $4 \times 4$ DCT band of QCIF format is 1584.
  - FW-SIR : 12672. It is determined based on the fact that each field contains a half number of pixels in QCIF resolution.
- **Conventional methods used as references**
  - Conventional SIG : the method in the literature [7] without the refinement.
  - M.B.Badem-SIR : SIR method in the literature [21].
  - R.Martins-SIR : SIR method in the literature [22].
  - J.Ascenso-SIR : SIR method in the literature [18].
- **Correlation noise modeling (see [25] for more details)**
  - BW-SIR : Correlation noise model at coefficient/frame level for the offline modeling.
  - FW-SIR : Correlation noise model at pixel level for the offline modeling.
- **Codec for key frames**
  - H.264/AVC JM reference codec 17.1 version, Table 1 and Table 2 show the QP of key frames corresponding to each $Q_i$ for BW-SIR and FW-SIR, respectively.
- **Test sequences**
  - BW-SIR : Foreman, Soccer, Stefan, and Coastguard.
  - FW-SIR : Foreman, Soccer, and Stefan.
- **Resolution**
  - QCIF format at 15 Hz.
- **Frames for each sequence**
  - 105 frames of each sequence are selected i.e. Foreman (110 ~ 214), Soccer (100 ~ 204), Stefan (194 ~ 298), and Coastguard (0 ~ 104). The selected frames for Foreman, Soccer and Stefan contain relatively high motion.
- **GOP Sizes**
  - 2, 4, and 8.
- **Bitrates and PSNR**
  - computed with the luminance component of the WZ frames.
- **Block size, search range and other parameter set**
  - Initial SIG (see [2] and [7] for more details) : forward motion estimation between key frames with $16 \times 16$ block size and ±32 pixel search range, and bidirectional motion
Table 1  QP of key frames for TDWZ codec evaluation.

|       | Q₁ | Q₂ | Q₃ | Q₄ | Q₅ | Q₆ | Q₇ | Q₈ |
|-------|----|----|----|----|----|----|----|----|
| Foreman | 40 | 39 | 38 | 34 | 32 | 29 | 25 |
| Soccer | 44 | 43 | 41 | 36 | 34 | 31 | 25 |
| Stefan | 40 | 39 | 38 | 34 | 32 | 29 | 25 |
| Coastguard | 38 | 37 | 37 | 34 | 33 | 31 | 30 | 36 |

Table 2  QP of key frames for PDWZ codec evaluation.

|       | Q₁ | Q₂ | Q₃ | Q₄ | Q₅ |
|-------|----|----|----|----|----|
| Foreman | 39 | 38 | 34 | 33 | 32 |
| Soccer | 43 | 43 | 36 | 35 | 34 |
| Stefan | 39 | 38 | 34 | 33 | 32 |

estimation with 8×8 block size and adaptive search range.
BW-SIR : 4×4 block size, search range of ±16 pixels and \( \alpha = 1 \) (for Foreman, Soccer, and Stefan sequence), \( \alpha = 5 \) (for Coastguard sequence)
FW-SIR : 4×4 block size, search range of ±8 pixels and \( \alpha = 3 \) (for Foreman, Soccer, and Stefan sequence)
R.Martins-SIR : 4×4 block size, search range of ±4 pixels and \( \mu = 100 \) (see [22] for more details)
M.B.Badem-SIR : 8×8 block size and search range of ±16 pixels
J.Ascenso-SIR : 4×4 block size, search range of ±16 pixels and \( \tau = 1.5 \) per pixel (see [18] for more details).

4.1 Performance Evaluation of BW-SIR Method

To evaluate the performance of the proposed BW-SIR method, the TDWZ codec explained in Sect. 2 is implemented and several experiments have been performed.

Figure 12 and Fig. 13 show the difference of visual image quality of the SI and the decoded frame between the conventional [7] and the proposed method. As shown in these figures, the SI generated by the conventional method reveals severe blocking artifacts because of the limited performance of the simple MCTI process. On the contrary, the refined SI of the proposed method shows gradual improvement of image quality as the frequency band goes higher. For example, Fig. 12(d), a refined SI generated after decoding the third AC coefficients shows better image quality than Fig. 12(c). This improvement in the SI also results in a better quality of the decoded WZ frame. Therefore, the reconstructed frame of the proposed method shows superior visual quality over that of the conventional SIG method.

Figure 14 shows the average PSNR of the refined SIs after decoding each band. As expected, it is observed that the objective visual quality of the SI is gradually improved as the frequency band goes higher, which leads to the reduction of the required parity bits compared with the conventional SIG method.

The RD performance (including only WZ frames) of the proposed method is compared with that of the conventional SIG method [7] and two conventional SIR methods [21], [22], and the results for four test sequences are shown in Figs. 15–18, respectively. Eight RD points are
Fig. 15  Comparison of RD performance (Foreman) : (a) GOP=2, (b) GOP=4, (c) GOP=8.

Fig. 16  Comparison of RD performance (Soccer) : (a) GOP=2, (b) GOP=4, (c) GOP=8.

Fig. 17  Comparison of RD performance (Stefan) : (a) GOP=2, (b) GOP=4, (c) GOP=8.
computed for each sequence. The improvement in RD performance is observed for all bitrates. It is also possible to observe that the extent of the improvement becomes large according to the increase of the bitrates. It comes from that more bits allocation for a WZ frame makes the proposed method generate more accurate SI.

Comparing the RD performances of each sequence, the Stefan sequence shows the largest RD improvements. At the last RD point, the RD gain is about 0.8 dB for GOP size 2 and 2.0 dB for GOP size 8 compared to the conventional SIG. Comparing the proposed BW-SIR with the R.Martins-SIR [22], the RD gain is about 0.4 dB for GOP size 2 and 0.9 dB for GOP size 8. The trends of RD plots for Soccer and Foreman sequences are similar to Stefan sequence; this means higher gains for the sequences with more complex motion, for the higher quantization indexes and for the longer GOPs. The RD gains observed for the Foreman and Soccer sequences can go up to 1.5 and 1.2 dB for GOP size 8, respectively, compared to the conventional SIG. And the RD gains are about 0.5 and 0.35 dB, respectively, compared to the R.Martins-SIR. The Coastguard sequence does not show such big improvements because the sequence does not contain complex motion over whole frames. The gains observed for GOP size 8 is about 0.6 dB for Coastguard compared to the conventional SI.

4.2 Performance Evaluation of FW-SIR Method

To evaluate the performance of the FW-SIR method, the PDWZ codec explained in Sect. 4 is implemented. Figure 19 shows the PSNRs of the SIs for each field. As observed, the objective quality of the bottom field SI is dramatically improved compared to the top field SI. The average improvements over 50 frames are 5.4 dB in the “Foreman” sequence, and 4.6 dB in the “Soccer” sequence. The PSNR increase in the SI will lead to an overall RD performance improvement of the codec.

The performance (including only WZ frames) of the proposed methods is compared with that of the conventional SIG method [7] and the conventional PDWZ-based SIR method [18]. Figures 20–22 present the RD performance for three test sequences including complex motion, namely Foreman, Soccer, and Stefan. Five RD points are computed for each sequence. There are two proposed methods; one method utilizes the field-level interleaving only (identified by “FW-SIR”), and the other method utilizes the field-level and bitplane-level interleaving simultaneously (identified by “FW+bit-plane SIR”). Compared with the conventional PDWZ-based SIR, the performance improvement of the proposed method that utilizes the field-level interleaving only is not large. However, it is possible to observe that the proposed method that utilizes both the field-level and bitplane-level interleaving shows the best performance for
Fig. 20  Comparison of RD performance (Foreman) : (a) GOP=2, (b) GOP=4, (c) GOP=8.

Fig. 21  Comparison of RD performance (Soccer) : (a) GOP=2, (b) GOP=4, (c) GOP=8.

Fig. 22  Comparison of RD performance (Stefan) : (a) GOP=2, (b) GOP=4, (c) GOP=8.
all bitrates. We can also observe that this improvement becomes larger with the increase of the bitrates.

5. Conclusions

In this paper, a progressive side information refinement framework for the WZ codec is proposed. Taking advantage of the already reconstructed data, the quality of the side information for the remaining data is gradually improved by the proposed SI refining process. Two specific methods, BW-SIR and FW-SIR, applying this framework are proposed for TDWZ and PDWZ codec, respectively. Simulation results show that the proposed methods improve the quality of the SI and rate-distortion performance compared to the conventional methods.

Acknowledgments

This work was supported by the Korea Research Foundation Grant Funded by the Korean Government (KRF-2008-313-D00705).

References

[1] Y. Vatis, S. Klomp, and J. Ostermann, “Enhanced reconstruction of the quantized transform coefficients for Wyner-Ziv coding,” Proc. IEEE Int. Conf. on Multimedia & Expo, pp.172–175, Beijing, China, July 2007.
[2] J. Ascenso, C. Brites, and F. Pereira, “Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding,” Proc. 5th EURASIP Conf. on Speech and Image Processing, Smolenic, Slovak Republic, July 2005.
[3] C. Brites, J. Ascenso, and F. Pereira, “Improving transform domain Wyner-Ziv video coding performance,” Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, vol.2, pp.525–528, Toulouse, France, May 2006.
[4] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” IEEE Trans. Inf. Theory, vol.19, no.4, pp.471–480, July 1973.
[5] A.D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” IEEE Trans. Inf. Theory, vol.22, no.1, pp.1–10, Jan. 1976.
[6] X. Huang and S. Forchhammer, “Improved virtual channel noise model for transform domain Wyner-Ziv video coding,” Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp.921–924, Taipei, Taiwan, April 2009.
[7] J. Ascenso, C. Brites, and F. Pereira, “Content adaptive Wyner-Ziv video coding driven by motion activity,” Proc. IEEE Int. Conf. on Image Processing, pp.605–608, Atlanta, USA, Oct. 2006.
[8] A. Aaron, E. Setton, and B. Girod, “Towards practical Wyner-Ziv coding of video,” Proc. IEEE Int. Conf. on Image Processing, vol.2, pp.869–872, Barcelona, Spain, Sept. 2003.
[9] S. Klomp, Y. Vatis, and J. Ostermann, “Side information interpo-lation with sub-pel motion compensation for Wyner-Ziv decoder,” Proc. Int. Conf. on Signal Processing and Multimedia Applications, Setubal, Portugal, Aug. 2006.
[10] S. Argypoulos, N. Thomos, N. Boulgouris, and M. Strintzis, “Adaptive frame interpolation for Wyner-Ziv video coding,” Proc. IEEE Int. Workshop on Multimedia Signal Processing, pp.159–162, Chania, Greece, Oct. 2007.
[11] R. Puri and K. Ranichandran, “PRISM: A ‘reversed’ multimedia coding paradigm,” Proc. IEEE Int. Conf. on Image Processing, vol.1, pp.617–620, Barcelona, Spain, Sept. 2003.
[12] D. Varodayan, D. Chen, M. Flierl, and B. Girod, “Wyner-Ziv coding of video with unsupervised motion vector learning,” Signal Process.: Image Commun., vol.23, no.5, pp.369–378, June 2008.
[13] X. Huang and S. Forchhammer, “Improved side information generation for distributed video coding,” Proc. IEEE Int. Workshop on Multimedia Signal Processing, pp.223–228, Cairns, Australia, Oct. 2008.
[14] Y.K. Chen, A. Vetro, H. Sun, and S.Y. Kung, “Frame-rate up-conversion using transmitted true motion vectors,” Proc. IEEE Int. Workshop on Multimedia Signal Processing, pp.622–627, California, USA, Dec. 1998.
[15] B.T. Choi, S.H. Lee, and S.J. Ko, “New frame rate up-conversion using bi-directional motion estimation,” IEEE Trans. Consum. Electron., vol.46, no 3, pp.603–609, Aug. 2000.
[16] A. Aaron, S. Rane, and B. Girod, “Wyner-Ziv video coding with hash-based motion compensation at the receiver,” Proc. IEEE Int. Conf. on Image Processing, vol.5, pp.3097–3100, Singapore, Oct. 2004.
[17] J. Ascenso and F. Pereira, “Adaptive hash-based side information exploitation for efficient Wyner-Ziv video coding,” Proc. IEEE Int. Conf. on Image Processing, vol.3, pp.29–32, Texas, USA, Sept. 2007.
[18] J. Ascenso, C. Brites, and F. Pereira, “Motion compensated refinement for low complexity pixel based distributed video coding,” Proc. IEEE Int. Conf. Advanced Video Signal Based Surveillance, pp.593–598, Como, Italy, Sept. 2005.
[19] A.B.B. Adikari, W.A.C. Fernando, H.K. Arachchi, and W.A.R.J. Weerakkody, “Sequential motion estimation using luminance and chrominance information for distributed video coding of Wyner-Ziv frames,” Electron. Lett., vol.42, no.7, pp.398–399, March 2006.
[20] W.A.R.J. Weerakkody, W.A.C. Fernando, J.L. Martinez, P. Cuenca, and F. Quiles, “An iterative refinement technique for side information generation in DVC,” Proc. IEEE Int. Conf. on Multimedia & Expo, pp.164–167, Beijing, China, July 2007.
[21] M.B. Badem, M. Mrak, and W.A.C. Fernando, “Side information refinement using motion estimation in DC domain for transform-based distributed video coding,” Electron. Lett., vol.44, no.16, pp.965–966, July 2008.
[22] R. Martins, C. Brites, J. Ascenso, and F. Pereira, “Refining side information for improved transform domain Wyner-Ziv video coding,” IEEE Trans. Circuits Syst. Video Technol., vol.19, no.9, pp.1327–1341, Sept. 2009.
[23] R. Martins, C. Brites, J. Ascenso, and F. Pereira, “Statistical motion learning for improved transform domain Wyner-Ziv video coding,” IET Image Process., vol.4, no.1, pp.28–41, 2010.
[24] D. Varodayan, A. Aaron, and B. Girod, “Rate-adaptive codes for distributed source coding,” EURASIP Signal Processing Journal, Special Section on Distributed Source Coding, vol.86, no.11, Nov. 2006.
[25] C. Brites and F. Pereira, “Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding,” Proc. IEEE Trans. Circuits Syst. Video Technol., vol.18, no.9, pp.1177–1190, Sept. 2008.
[26] D. Kubasov, J. Nayak, and C. Guillenmot, “Optimal reconstruction in Wyner-Ziv video coding with multiple side information,” Proc. IEEE Int. Workshop on Multimedia Signal Processing, pp.183–186, Chania, Greece, Oct. 2007.
Chan-Hee Han received his MS degree in Multimedia Engineering from Hanbat National University in 2008. Since March 2008, he has been pursuing a Ph.D. degree at the Graduate School of Information and Communications, Hanbat National University, Korea. His research interests include video codec and computer vision.

Si-Woong Lee received his Ph.D. degree in electrical and electronics engineering from Korea Advanced Institute of Science and Technology (KAIST) in 1997. From 1995 to 2000, he was with Samsung Electronics Co., Ltd., where he was engaged in digital IC design as a senior engineer. Since April 2000, he has been a professor with Division of Information, Communication and Computer Engineering of Hanbat National University, Korea. His research interests include digital video compression, image processing and computer vision.

Hamid GholamHosseini received his M.Sc. in Electrical and Electronic Engineering, in 1985 (Tehran University, Iran) and his Ph.D. in Biomedical Engineering in 2002 (Flinders University, Australia). He is presently a senior lecturer at School of Engineering and a research group leader at Engineering Research and Innovation Cluster (ERIC), Auckland University of Technology, Auckland, New Zealand. His research interests include biomedical signal and image processing, electronic nose, embedded and reconfigurable systems, and remote monitoring systems.

Yun-Ho Ko received his Ph.D. degree in electrical and electronics engineering from Korea Advanced Institute of Science and Technology (KAIST) in 2002. From 2002 to 2004, he was with Hynix Semiconductor Co., Ltd., where he was engaged in digital IC design as a senior engineer. Since October 2004, he has been an assistant professor with Department of Mechatronics Engineering of Chungnam National University, Korea. His research interests include digital video compression, image processing and computer vision.