

SELECTIVE DATA PRUNING BASED DISTRIBUTED VIDEO CODING WITH MODIFIED HIGH-ORDER EDGE-DIRECTED INTERPOLATION

Ha Vu Le[†], Kazuma Shinoda[†], Madoka Hasegawa[†], Shigeo Kato[†], and Yuichi Tanaka[‡]

[†]Graduate School of Engineering, Utsunomiya University, Utsunomiya, Tochigi, 321-8585 Japan

[‡]Graduate School of BASE, Tokyo University of Agriculture and Technology, Koganei, Tokyo, 184-8588 Japan

Email: [†]{11levu@mclaren.is, shinoda@is, madoka@is, kato@is}.utsunomiya-u.ac.jp, [‡]ytink@cc.tuat.ac.jp

ABSTRACT

In distributed video coding (DVC), a low resolution video sequence is usually generated with spatial and/or temporal downsampling at an encoder. At a decoder side, interpolation is performed and the interpolated pixels are further refined by using an error-correcting code such as Turbo codes or LDPC. In our previous work, we proposed a spatial domain DVC which uses a line-based downsampling method called selective data pruning (SDP). SDP is used for reducing the spatial frame size before compression. After decoding, received frames are decoded and interpolated back to their original size by high-order edge-directed interpolation (HEDI). In this paper, we propose a modified high-order edge-directed interpolation (M-HEDI) to avoid interpolation artifacts. Experimental results show that the interpolated frames by using M-HEDI have higher quality than those using HEDI. The proposed method outperforms conventional DVCs and the video coding standard H.264/AVC in majority of all the test video sequences.

Index Terms— Video coding, data pruning, edge-directed interpolation, distributed video coding.

1. INTRODUCTION

Distributed source coding (DSC) is based on two information theory results: the theorems by Slepian and Wolf [1] and Wyner and Ziv [2] for lossless and lossy codings of correlated source, respectively. Recently, practical DSC schemes have received attention in efforts such as distributed video coding (DVC) with reversed complexity, improving error resilience, efficient multi-view coding for distributed cameras, and flexible decoding capability. In the conventional video coding standards, an encoder has high complexity mainly due to motion estimation for finding the best matching block in a reference frame whereas a decoder has low complexity. In contrast, DVC enables to reverse the complexity of codecs, where the decoder requires more computational effort than the encoder. Even though, DVC algorithms do not outperform the conventional video coding schemes at present, but it is regarded as a suitable codec for power-constrained (mobile or hand-held) devices.

A particular application of DVC is transform domain Wyner-Ziv (TDWZ) codec [3,4]. In this method, key frames are encoded by the conventional intra mode while the Wyner-Ziv (WZ) frames (non-key frames) are transformed and encoded using error correcting codes. At the decoder, the WZ frames are estimated from the received key frames and corrected in the transform domain Wyner-Ziv decoder by using transmitted parity bits.

On the other hand, at the image coding standpoint, there exist several methods of decimation-then-compression (DTC) approach

[5, 6]. These methods are performed as follows. First, insignificant regions are discarded from an input image, and then the remaining image is compressed by some image coding standards. The compressed small image and the side information, i.e., positions of discarded pixels, are transmitted. On the receiver side, the small image and the side information are synthesized to reconstruct an image of the same size and structure as the original image.

From the idea of DTC, Vö et al. proposed a video coding scheme based on selective data pruning (SDP) and high order edge-directed interpolation (HEDI) to reduce the bitrate while keeping high quality of the reconstructed video [7]. In this scheme, SDP is applied to optimally prune the original frames to a smaller size by adaptively dropping rows and/or columns prior to encoding. Besides, the HOEI is applied to reconstruct the decoded frames to their original size.

In our previous work, we proposed a video coding scheme called SDP-DVC [8,9] to improve SDP-based video coding performances by using an error-correcting system of DVC. In particular, we applied TDWZ coding for residual signals which are differences between the pruned lines by SDP and their interpolated ones. The method outperforms the conventional DVC for video sequences containing a large number of motions. However, artifacts are still visible in the interpolated (and corrected) frames.

In this paper, we propose a SDP-DVC with the modified high-order edge-directed interpolation (M-HEDI) which uses not only the pixels of current frames but also available pixels in the previous key frame to take into account the temporal correlation between frames for the interpolation at the decoder.

The rest of this paper is organized as follows. Section 2 briefly reviews new edge-directed interpolation (NEDI) [11] and the further improved edge-directed interpolation (FEDI) [10]. In Section 3, we present our proposed algorithm. Section 4 shows experimental results. Finally, section 5 concludes the paper.

2. RELATED WORK

2.1. New Edge-Directed Interpolation

The NEDI method [11] is based on the idea of the geometric duality: the covariance of the high-resolution image is estimated from its low-resolution counterpart. The model of NEDI is shown in Fig.1(a). In Fig.1(a), the blue dots represent the low-resolution pixels of low-resolution image. The original NEDI involves two steps: (i) interpolation with four symmetrical neighbors to obtain the predicted value of the center unknown pixel (orange dot), and (ii) fourth-order interpolation of the remaining pixel. This method is originally used for interpolating an image to double the size of rows and columns.

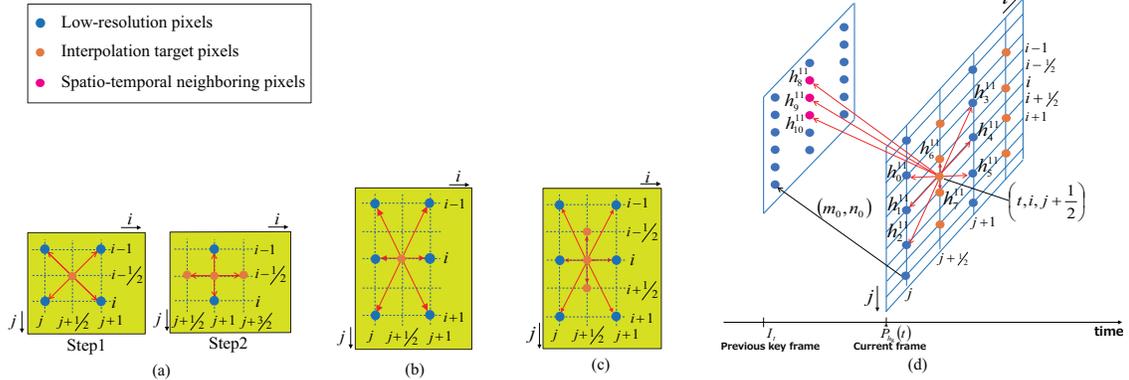


Fig. 1. Illustrative example of various interpolation methods. (a) NEDI. (b) HEDI-6. (c) FEDI. (d) M-HEDI.

However, in SDP-based video coding method, the aspect ratio of input video is usually changed by SDP. If some columns are pruned, only the horizontal resolution becomes low.

To address this issue, HEDI was proposed to utilize the geometric duality after SDP [7]. For the single frame-based interpolation, the sixth-order edge-directed interpolation (HEDI-6) is developed to interpolating rows or columns. An example of HEDI-6 is shown in Fig.1(b). It uses six neighboring pixels to interpolate the target pixel (orange dot).

2.2. Further Improved Edge-Directed Interpolation

FEDI is proposed to improve the performance of NEDI. The model of FEDI is shown in Fig.1(c). Its first interpolation step is the same as NEDI, which uses the fourth-order linear prediction to interpolate the unknown pixels.

In the second step, two interpolated results obtained from the first step (orange dots) are used together with six neighboring pixels (blue dots) to obtain the interpolation target pixel (the center orange dot). This structure can remove most of the worst estimation and sharpen the edge of the image by making a full use of the local relative information in the low-resolution image.

3. SDP-DVC WITH MODIFIED HIGH-ORDER EDGE-DIRECTED INTERPOLATION

3.1. Modified High-Order Edge-Directed Interpolation

For SDP-based video coding, HEDI (or FEDI) can be applied to interpolate each frame in a video sequence. However, artifacts and flickering effects are visible in the interpolated frames since the temporal correlation between frames is ignored. In order to improve FEDI, we propose a new multiframe-based interpolation called modified high-order edge-directed interpolation (M-HEDI). M-HEDI uses not only the pixels of the current frame but also available pixels in the previous key frame to interpolate the pixels which pruned by SDP process. Indeed, it can be considered as an extension of FEDI with the inspiration from HEDI-9 [7].

Let $\hat{P}_{h^s}(t)$ be the t -th low-resolution frame which need to be interpolated. Assume that I_t is the key frame within the group of pictures (GOP). The model of M-HEDI is shown in Fig.1(d). The interpolation process of M-HEDI is performed as follows.

1. Similarly to FEDI, the fourth-order linear prediction is used to interpolate the unknown pixels (orange dots).
2. Motion estimation (ME) and motion compensation (MC) are applied to align the current block of pixels in $\hat{P}_{h^s}(t)$ to its matching block in I_t . Then, two interpolated pixels obtained from the first step (orange dots) and six neighboring pixels (blue dots) in $\hat{P}_{h^s}(t)$ are used together with three pixels (pink dots) in the matching block in I_t to interpolate one high-resolution pixel (green dots) in the current frame.

It is formalized as follows:

$$\begin{aligned}
 \hat{P}_{h^s}(t, i, j + \frac{1}{2}) &= \sum_{k=-1}^1 \sum_{l=0}^1 h_{3l+k+1}^{11} \times \hat{P}_{h^s}(t, i+k, j+l) \\
 &+ \sum_{k=0}^1 \sum_{l=0}^1 h_{6+k}^{11} \times \hat{P}_{h^s}(t, i+k-\frac{1}{2}, j+\frac{1}{2}) \\
 &+ \sum_{k=-1}^1 h_{9+k}^{11} \\
 &\times I_t(i+k+m, j+1+n) \\
 &= \mathbf{h}^{11} \cdot \mathbf{P}_{h^{11}}
 \end{aligned} \tag{1}$$

where $\hat{P}_{h^s}(t, i, j)$ is (i, j) -th pixel in $\hat{P}_{h^s}(t)$, \mathbf{h}^{11} is the vector of eleventh-order model parameters [7], $\mathbf{P}_{h^{11}}$ is the vector of 8-spatial neighboring pixels and 3-spatio-temporal neighboring pixels, and (m, n) is the motion vector of the current block. The first and second terms in equation (1) represent the spatial pixels in $\hat{P}_{h^s}(t)$ and the third term represents the spatio-temporal pixels in I_t . The filter-like interpolation of M-HEDI helps reducing the artifact and flickering effect of using only the single frame-based interpolation.

3.2. SDP-DVC based on M-HEDI

In this section, we present a new SDP-DVC using M-HEDI to improve the performance of our previous SDP-DVC scheme [8,9]. Different from the previous work in [8,9], we use FEDI for interpolation process at the encoder and M-HEDI at the decoder. The architecture of the proposed SDP-DVC is shown in Fig. 2.

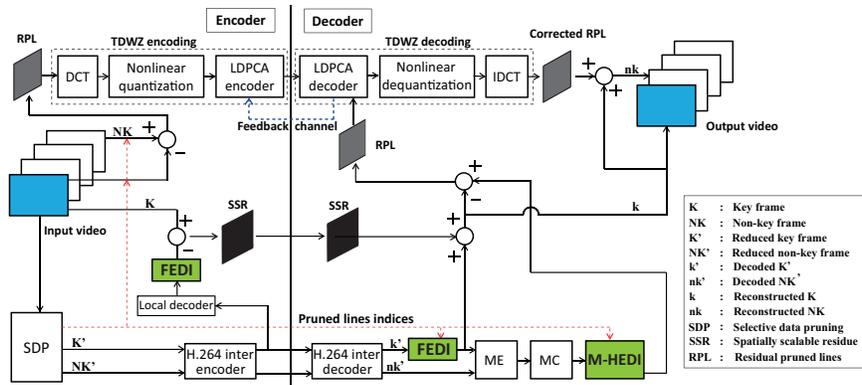


Fig. 2. Architecture of the proposed SDP-DVC.

3.2.1. Encoder Architecture

Firstly, all frames are resized by SDP and the resized frames are encoded with H.264/AVC inter mode. For all frames within a GOP, the downsampling positions by SDP are the same.

In the next step, resized key frames are decoded by a local decoder and interpolated back to their original size by using FEDI. Then, the full resolution residue between the interpolated key frames and the corresponding original frames, which are called the spatially scalable residue (SSR) hereafter, are also encoded with H.264/AVC inter mode.

For non-key frames, lines of pixels which locate at the pruned position are extracted by using pruned position indices. After that, residual pruned lines (RPLs) are obtained by taking the difference between the pruned lines of non-key frames and those of the preceding key frames. Here, only the errors of RPLs are corrected to improve the accuracy of error-correcting process. RPLs are encoded by TDWZ coder and will be used for correcting interpolated lines at the decoder. During the TDWZ encoding, the LDPC-accumulated (LDPCA) codec [12] is used as the error-correcting code. The number of WZ frames and the number of pruned lines may be varied dynamically based on the complexity reduction target.

For RPLs, every 16 one-dimensional signals are rearranged into a 4×4 matrix by zigzag scanning in order to implement two-dimensional discrete cosine transform (DCT). We also applied nonlinear quantization [13] for the DCT coefficients of RPLs represented as follows:

$$y = \beta \text{sgn}(x) \left(1 - e^{-\frac{c|x|}{\beta}}\right) \quad (2)$$

$$\hat{x} = -\frac{\beta}{c} \text{sgn}(y) \ln\left(1 - \frac{|y|}{\beta}\right) \quad (3)$$

where x , y , and \hat{x} represent input, quantized, and reconstructed RPLs signals, respectively. $\text{sgn}(\cdot)$ is a function that returns the sign of the arguments. The arbitrary parameters c and β are constants that represent the level of nonlinearity.

3.2.2. Decoder Architecture

At the decoder, the reconstruction process is performed as follows. First, decoded low-resolution key frames are interpolated to the original size by FEDI. The interpolated key frames are incorporated with SSR to reconstruct full resolution key frames. For the non-key frames, block-based ME/MC is applied to find the most similar block in the previous full-resolution key frame. If the sum of absolute difference (SAD) between the current block and its matching

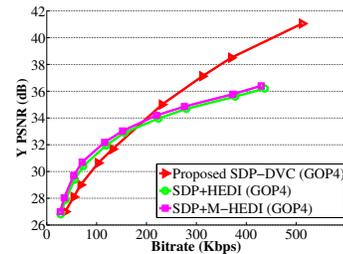


Fig. 3. Performance comparison of SDP-based methods for *Foreman*.

block is greater than a threshold θ , the three spatio-temporal pixels of M-HEDI are discarded since the pixels in the matched block are not related. Then, M-HEDI is applied to interpolate the non-key frames. Moreover, for non-key frames, the temporal residue is calculated to obtain the side information for TDWZ decoder. After TDWZ decoding, the reconstructed non-key frames are obtained by adding the corrected residue to previous reconstructed key frames.

4. EXPERIMENTAL RESULTS

In this section, the experimental results are shown to validate the effectiveness of our proposed method. We used luminance (Y) signal of 150 frames of popular video sequences, *Foreman*, *Hall Monitor*, and *Soccer* with QCIF size and 15fps. The number of pruned lines by SDP in row and column are fixed to 48, which means about a half of pixels in each frame are removed. The PSNR and bitrates are averaged over the sequences. Besides, we used the same condition as that of [8] for pruned indices and nonlinear quantization parameters. As the core codec used in the encoder, we used IPPP coding of H.264/AVC.

4.1. Effect of using M-HEDI

PSNRs of interpolated frames in the decoder by using two methods of HEDI (HEDI-6, HEDI-9) and M-HEDI without error correction are shown in Table 1. The interpolated frames using M-HEDI have higher PSNR for all frames than the conventional methods. Especially, there are highly effective for slow motion video sequence such as *Hall Monitor*.

Moreover, Fig. 3 shows the performance comparison among the proposed SDP-DVC, the SDP-based video coding with HEDI alone (SDP+HEDI) and the SDP-based video coding with M-HEDI alone (SDP+M-HEDI) for *Foreman*. The experimental result shows that SDP+M-HEDI has better performance than SDP+HEDI. Clearly

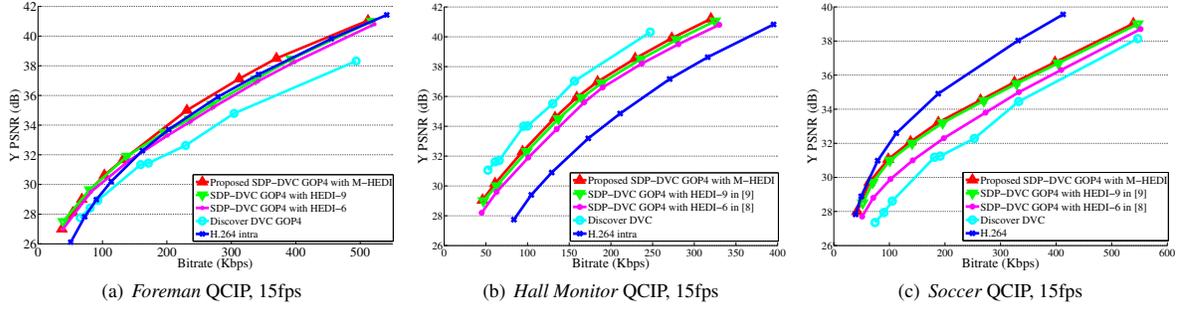


Fig. 4. Rate-distortion performance of various coding methods.

Table 1. Mean PSNR comparison (in dB)

Sequences (GOP4)	HEDI-6	HEDI-9	M-HEDI
Foreman	34.80	35.66	35.77
Hall Monitor	29.44	29.73	30.34
Soccer	33.81	34.25	34.30

from middle to high bitrates, the proposed SDP-DVC gains significant PSNR improvements. It is worth noting that SDP+HEDI is an option in the SDP-DVC system since it can be realized with the low-resolution bitstream and pruned indices. Naturally the performance intersection can be varied according to video sequence characteristics.

4.2. SDP-DVC Performance

The rate-distortion (R-D) performance of the proposed SDP-DVC is compared with the original SDP-DVC [8], DISCOVER DVC codec [14] with GOP of 4, our previous work [9], and H.264/AVC intra coding [15].

Fig. 4 illustrates the R-D curves of various video coding methods. Experimental results show that the proposed method outperforms the original SDP-DVC and SDP-DVC with HEDI-9 in all cases. It also outperforms DISCOVER for fast moving sequences, such as *Foreman* and *Soccer*. Moreover, it also shows better performance than H.264/AVC intra coding for *Hall Monitor* and *Foreman* sequence. Note that, SDP-DVC performance gradually reaches to H.264/AVC by using M-HEDI.

In our framework, the multiframe-based interpolation is effective for reducing the artifacts which are occurred by SDP. Fig. 5 shows the decoded frames of *Foreman* obtained by DISCOVER DVC, the previous works in [8, 9] and the proposed SDP-DVC. It can be observed that the proposed SDP-DVC shows better image quality than those of the conventional methods.

5. CONCLUSIONS

In this paper, SDP-DVC using multiframe-based interpolation M-HEDI was proposed. The R-D performance of the proposed codec is better than the conventional ones with relatively many motions. It also has better performance than H.264/AVC intra coding for *Hall Monitor* and *Foreman* sequences. There is a room for improvements of the interpolation. Future work also includes replacing SDP with other sophisticated resizing methods such as fast content aware re-targeting (CAR) [16–18].

6. ACKNOWLEDGEMENT

This work was supported in part by JSPS KAKENHI Grant Number 25420355.



(a) DISCOVER DVC (425Kbps, 37.00dB)



(b) Original SDP-DVC (SDP-DVC with HEDI-6) [8] (398Kbps, 38.25dB)



(c) SDP-DVC with HEDI-9 [9] (393Kbps, 38.45dB)



(d) Proposed SDP-DVC (370Kbps, 38.50dB)

Fig. 5. Performance comparison for 48th frame of *Foreman*.

7. REFERENCES

- [1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471-480, 1973.
- [2] A. 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, 1976.
- [3] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in *Proc. SPIE VCIP'04*, pp. 520-528, 2004.
- [4] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner-Ziv video coding performance," in *Proc. ICASSP'06*, 2006.
- [5] S. D. Rane, G. Sapiro, and M. Bertalmio, "Structure and texture filling-in of missing image blocks in wireless transmission and compression applications," *IEEE Trans. Image Process.*, vol. 12, no. 3, pp. 296-303, 2003.
- [6] D. Liu, X. Sun, F. Wu, S. Li, and Y. Q. Zhang, "Image compression with edge-based inpainting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 10, pp. 1273-1287, 2007.
- [7] D. T. Võ, J. Solé, P. Yin, C. Gomila, and T. Q. Nguyen, "Selective data pruning-based compression using high-order edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 399-409, 2010.
- [8] T. T. Phan, Y. Tanaka, M. Hasegawa, S. Kato, "Mixed-Resolution Wyner-Ziv Video Coding Based on Selective Data Pruning," *MMSP2011*, pp. 2-6, Hangzhou, China, Oct. 17-19, 2011.
- [9] H. V. Le, T. T. Phan, Y. Tanaka, M. Hasegawa, and S. Kato, "Distributed video coding based on selective data pruning and high-order edge-directed interpolation," in *IIEEJ Image Electronics and Visual Computing Workshop 2012 (IEVC2012)*, Kuching, Malaysia, Nov. 2012.
- [10] Wong Chi Shing and Wan-Chi Siu, "Further Improved Edge-Directed interpolation and Fast EDI for SDTV to HDTV conversion", *The 18th European Signal Processing Conference (EUSIPCO'10)*, August 2010, Aalborg Denmark.
- [11] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521-1527, 2001.
- [12] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Processing*, vol. 86, no. 11, pp. 3123-3130, 2006.
- [13] M. B. Badem, R. Weerakkody, A. Fernando, and A. M. Kondoz, "Design of a non-linear quantizer for transform domain DVC," *IEICE Trans. on Fundamentals*, vol. 92, no. 3, pp. 847-852, 2009.
- [14] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," in *Proc. 26th Picture Coding Symposium*, 2007.
- [15] T. Weigand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560-576, 2003.
- [16] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, 2007.
- [17] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, 2008.
- [18] Y. Tanaka, M. Hasegawa, and S. Kato, "Seam carving with rate-dependent seam path information," in *Proc. ICASSP'11*, pp.1449-1452, May 2011.