Adaptive Motion Vector Smoothing for Improving Side Information in Distributed Video Coding

Jun Guo* and Joohee Kim*

Abstract—In this paper, an adaptive motion vector smoothing scheme based on weighted vector median filtering is proposed in order to eliminate the motion outliers more effectively for improving the quality of side information in frame-based distributed video coding. We use a simple motion vector outlier reliability measure for each block in a motion compensated interpolated frame and apply weighted vector median filtering only to the blocks with unreliable motion vectors. Simulation results show that the proposed adaptive motion vector smoothing algorithm improves the quality of the side information significantly while maintaining low complexity at the encoder in frame-based distributed video coding.

Keywords—Distributed Video Coding, Wyner-Ziv Coding, Spatial Motion Vector Smoothing, Weighted Vector Median Filter, Side Information

1. INTRODUCTION

Ever since the Slepian-Wolf theorem [1] and the Wyner-Ziv (WZ) theorem [2] have been applied to Distributed Video Coding (DVC) [3], several practical coders have been proposed [4-6]. The Stanford version is a frame-based architecture using turbo codes and a feedback channel to perform rate control at the decoder. The Power-efficient Robust hIgh-compress Syndrome-based Multimedia codec (PRISM) works at block level and does not require a feedback channel. The DIStributed COding for Video sERvices (DISCOVER) codec is also a frame-based architecture, but improves the Stanford codec in Side Information (SI) generation, reconstruction, channel coding and so on. In [7], the rate-distortion performance of these codecs has been discussed.

In frame-based DVC, a video sequence is classified into key frames and WZ frames at the encoder. The key frames are encoded by using a conventional intra coding method (e.g., H.264/AVC intra mode [8]), while the frames in between key frames are WZ encoded. The WZ frames are divided into 4×4 blocks and each block is transformed, quantized, and each bit-plane of the quantization indices is encoded using a channel code. At the decoder, key frames are decoded and used to create SI. Then the parity bits sent from the encoder are used to refine the SI. The quality of SI has a significant effect on the Rate-Distortion (RD) performance of DVC because the better the quality of SI, the fewer bits will be required to refine the WZ frame. In frame-based DVC, various Motion-Compensated Frame Interpolation (MCFI) techniques have been proposed to generate SI from the previous and next key frames [9]. Most of the MCFI me-

Manuscript received August 4, 2010; accepted August 31, 2010.
Corresponding Author: Jun Guo
* Dept. of Electrical and Computer Engineering, Illinois Institute of Technology, United States (jguo19@iit.edu, joohee@ece.iit.edu)

thods employ a block matching algorithm for motion estimation. However, when there is high or asymmetric motion, it is hard to find the true motion field and MCFI suffers from annoying visual artifacts. To improve the performance of MCFI in SI generation, Spatial Motion Vector Smoothing (SMVS) algorithms that remove Motion Vector (MV) outliers have been proposed [10]. The SMVS method proposed in [10] uses Weighted Vector Median Filtering (WVMF) to correct irregular MVs, but this technique will only be effective at removing isolated false MVs in smooth areas. In general, the assumption that the MV field is smooth is not always true in that a video frame may contain complex motion, especially on the motion boundaries. In other words, if the SMVS solution is performed on motion boundaries or other areas in which MVs are complex, the MVs of SI would not be refined. Instead, the filtered MVs may cause SI to be corrupted by ghost artifacts or some other undesirable effects.

In this paper, an adaptive motion vector smoothing algorithm using a WVMF is proposed in order to improve the quality of SI in frame-based DVC. Rather than applying the filter to all MVs, the proposed algorithm adaptively selects the candidate MVs according to the Sum of Absolute Difference (SAD) value of each corresponding block. If the block SAD is smaller than a threshold, its MV is left unchanged. Otherwise, the MV is selected as a candidate for MV smoothing. Then, a 3×3 WVMF is performed to correct the motion outliers. Simulation results show that the SI generated by the proposed algorithm achieves significantly better quality compared to the SMVS based method proposed in [10], while maintaining low computational complexity of the encoder. The proposed method performs well for various video sequences with different motion types.

The rest of this paper is organized as follows. First, SMVS is reviewed in Section 2. In Section 3, the proposed method is presented in detail. Then, simulation results and analyses are presented in Section 4. Conclusions are given in Section 5.

2. SPATIAL MOTION VECTOR SMOOTHING

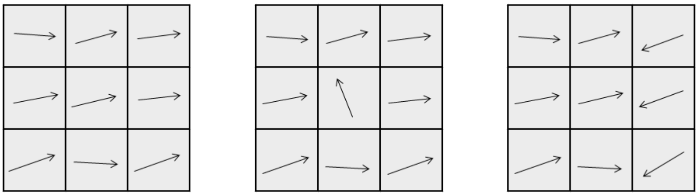
The SMVS method proposed in [10] employs a WVMF defined by:

$$\sum_{j=1}^{N} \omega_j \, \| \vec{\upsilon}_{wvmf} - \vec{\upsilon}_j \|_L \le \sum_{j=1}^{N} \omega_j \, \| \vec{\upsilon}_i - \vec{\upsilon}_j \|_L \qquad (1)$$

where $\vec{\upsilon}_1, \ldots, \vec{\upsilon}_N$ are the MVs of the current block in the previously interpolated frame and the corresponding nearest neighboring blocks. The vector $\vec{\upsilon}_{wvmf}$ is the selected MV that minimizes the sum of distances to the other $N-1$ neighboring vectors. The weights $\omega_1, \cdots, \omega_N$ are chosen according to the prediction error defined by:

$$\omega_i = \frac{MSE(\vec{\upsilon}_c, B)}{MSE(\vec{\upsilon}_i, B)} \qquad \text{i = 1, 2, …, N} \qquad (2)$$

Where $\vec{\upsilon}_c$ represents the candidate MV for the current block B to be smoothed. The Mean Squared Error (MSE) is defined by:



(a) (b) (c)

Fig. 1. (a) smooth region, (b) motion outlier, (c) motion boundary

(a) (b)

Fig. 2. The 2nd frame of Foreman sequence (30Hz, QCIF): (a) SI without SMVS (PSNR: 31.27dB). (b) SI with the SMVS method (PSNR: 31.3dB)

$$MSE_B(\vec{\upsilon}) = \frac{\sum_{m,n\in B}[I_{k+1}(m,n) - I_{k-1}(m+\vec{\upsilon}_x, n+\vec{\upsilon}_y)]}{BSize^2} \tag{3}$$

where I_{k+1} and I_{k-1} are the next and the previous key frames, respectively. m,n are the pixel coordinates in the current block. $\vec{\upsilon}_x$ and $\vec{\upsilon}_y$ are the horizontal and vertical components of $\vec{\upsilon}$, respectively, and $BSize$ is the block size.

By using this algorithm, in smooth areas, some motion outliers can be removed and restored effectively as in Fig. 1 (b). However, if the algorithm is applied in the areas with complex motion as in Fig. 1 (c), it turns out that ghost artifacts occur in the motion boundaries as shown in Fig. 2 (b).

3. PROPOSED ADAPTIVE MOTION VECTOR SMOOTHING METHOD

In order to address the problems of existing SMVS for SI generation, we propose an adaptive MV smoothing method. For adaptive MV filtering, MV reliability information is necessary. In [13], an MV reliability classification method that classifies an MV into three different reliability levels based on the reconstructed prediction error of each block has been proposed.

Since the original signal for the SI is not available at the decoder in DVC, we use the block SAD defined in equation (4) as the MV reliability measure.

$$SAD_B(\vec{\upsilon}) = \sum_{m,n\in B}[I_{k+1}(m,n) - I_{k-1}(m+\vec{\upsilon}_x, n+\vec{\upsilon}_y)] \tag{4}$$

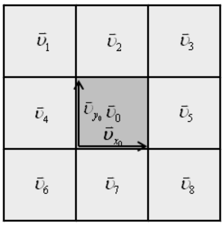


Fig. 3. A 3×3 weighted vector median filter

where $I_{k+1}, I_{k-1,} \vec{\upsilon}_x, \vec{\upsilon}_y, m, n$ are defined in the same way as in equation (3). In the proposed scheme, MVs are found in accordance with the smallest block SAD in the search range. We assume that the MV of a 16×16 block with large SAD is unreliable. The proposed WVMF is only applied to the blocks with unreliable MVs. Specifically, if the block SAD of a block is larger than a pre-defined threshold T, then WVMF is applied to this block. Otherwise, MV smoothing is not performed. The threshold for adaptive MV smoothing T is defined as the average of the block SAD values in a frame as in equation (5), where N is the number of blocks in a frame.

$$T = \frac{\sum_{i=1}^{N} SAD_B(\vec{\upsilon}_i)}{N} \tag{5}$$

A 3×3 WVMF is shown in Fig. 3, where $\vec{\upsilon}_0$ is the candidate MV for MV smoothing, $\vec{\upsilon}_1, \cdots, \vec{\upsilon}_8$ are its eight neighboring MVs, and $\vec{\upsilon}_{x_0}$ and $\vec{\upsilon}_{y_0}$ are the horizontal and vertical components of $\vec{\upsilon}_0$, respectively.

The weights are chosen in the same manner as in equation (1). That is, if

$$\sum_{j=1}^{8} \omega_j \,\|\, \vec{\upsilon}_i - \vec{\upsilon}_j \,\|_L \le \sum_{j=1}^{8} \omega_j \,\|\, \vec{\upsilon}_0 - \vec{\upsilon}_j \,\|_L ,$$

then we replace $\vec{\upsilon}_0$ with one of its neighboring MVs that has the smallest weighted L-norm value (left hand side of the previous inequality), $\vec{\upsilon}_i$.

4. RESULTS

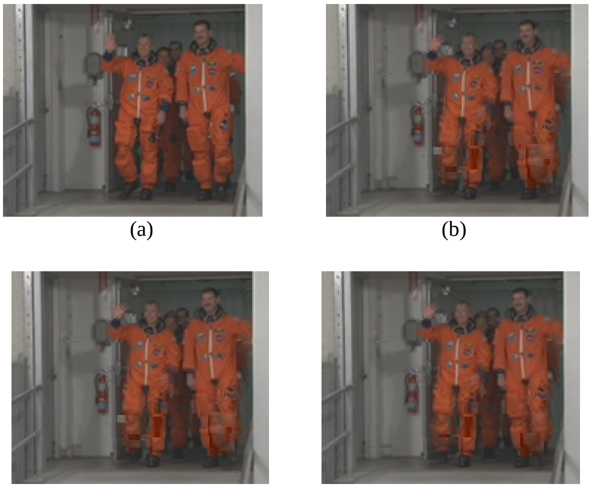
In our experiments, four Quarter Common Intermediate Format (QCIF: 176×144 pixels) sequences: Foreman, Crew, City and Soccer were used to compare the proposed adaptive MV smoothing method with the SMVS algorithm in [10]. The four video sequences were selected in order to test the performance of the proposed algorithm for various input video characteristics. Specifically, the Soccer sequence is a high motion sequence and the Foreman and Crew sequences contain a medium degree of motion. The City sequence contains a lot of high frequency components with a medium degree of motion. We implemented a frame-based transform-domain DVC codec and the Group Of Pictures (GOP) size was set to 2. As a result, the odd frames in the sequence are the key frames and the even frames are WZ frames. The key frames are intra coded using H.264/AVC [8]. For the WZ frames, the SI is generated by using Forward Motion Estimation (FME), Bidirectional Motion Estimation (BiME), and Bidirectional Motion Compensation (BiMC). For FME and BiME, a full search block matching algorithm was used. A 3×3 vector median filter was used in both the SMVS algorithm and the proposed approach.

The SI of frame 2 of the Foreman sequence generated by using the proposed method is shown in Fig. 4. It can be seen that the ghost artifacts present in Fig. 2 (b) have been removed.

In Figs. 5-8, the quality of the SI generated by the two methods has been evaluated for four different video sequences. Fig. 5 shows that the blocking artifacts around the lips have been removed by both the SMVS algorithm [10] and the proposed algorithm. However, the reconstructed image obtained by the proposed method is closer to the original frame. In Fig. 6, with the help of the proposed solution, the blocking artifacts around the legs of the crew have been removed, while the SI generated with the SMVS algorithm does not solve the problem. For the City sequence in Fig. 7, the proposed approach provides a high SI quality, while the SMVS method causes ghost artifacts around the top part of the building. In Fig. 8, ghost artifacts around the soccer players in the SI generated with the SMVS method are obvious. However, the SI obtained with the proposed algorithm does not show the artifacts.

Fig. 4. SI of frame 2 of Foreman sequence (30Hz, QCIF) using the proposed approach (PNSR: 31.30dB)

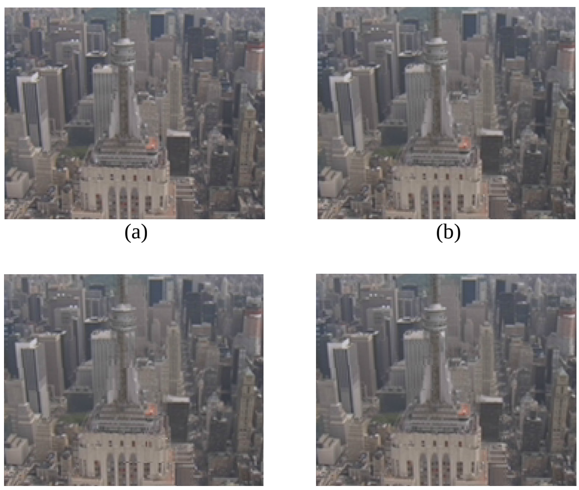
Fig. 5. Visual results comparison for Forman (30Hz, QCIF): (a) Original Frame, (b) SI without SMVS (PSNR: 29.40dB), (c) SI using the SMVS algorithm (PSNR: 29.57dB), (d) SI using the proposed method (PSNR: 29.70dB)



(a) (b)

(c) (d)

Fig. 6. Visual results comparison for Crew (15Hz, QCIF): (a) Original Frame, (b) SI without SMVS (PSNR: 31.43dB), (c) SI using the SMVS algorithm (PSNR: 31.43dB), (d) SI using the proposed method (PSNR: 32.39dB)

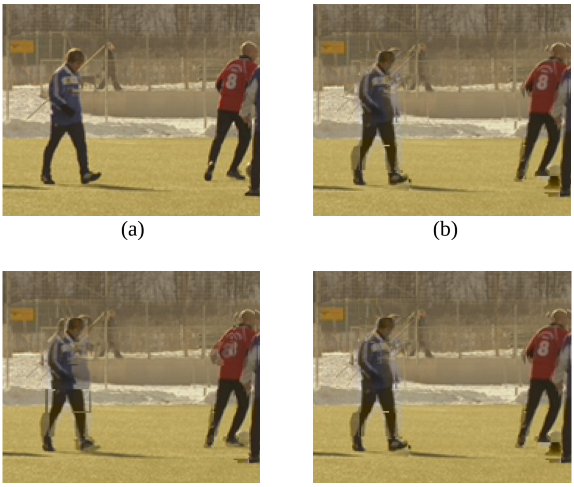


(a) (b)

(c) (d)

Fig. 7. Visual results comparison for City (15Hz QCIF): (a) Original Frame, (b) SI without SMVS (PSNR: 29.08dB), (c) SI using the SMVS algorithm (PSNR: 28.69dB), (d) SI using the proposed method (PSNR: 29.08dB)

To further evaluate the effectiveness of the proposed algorithm, 50 frames of four different video sequences have been encoded to evaluate the objective quality of SI generated by the different methods. In Table 1, the average Peak Signal-to-Noise Ratio (PSNR) of SI is generated by different methods. As seen from the results, the proposed method increases the PSNR of SI slightly.



(a) (b) (c) (d)

Fig. 8. Visual results comparison for Soccer (15Hz QCIF): (a) Original Frame, (b) SI without SMVS (PSNR: 24.65dB), (c) SI using the SMVS algorithm (PSNR: 24.70dB), (d) SI using the proposed method (PSNR: 24.81dB)

Table 1. Average PSNR of SI Generated by different methods

Sequences	SI without SMVS (dB)	SI generated by the SMVS (dB)	SI with the proposed method (dB)
oreman	30.85	30.91	30.92
Crew	28.97	29.00	29.13
City	29.71	29.70	29.75
Soccer	23.28	23.21	23.28

5. CONCLUSION

In this paper, we propose an adaptive motion smoothing algorithm to improve the quality of the SI in frame-based DVC. We define a simple MV reliability measure based on a block SAD and apply WVMF only for blocks with unreliable MV. The simulation results show that the proposed method is very effective in eliminating motion outliers in SI, while maintaining low complexity of the encoder. And the proposed algorithm outperforms existing techniques for various types of video sequences. In the future, we will improve our MV reliability classification method so that MV outliers can be identified more accurately and effectively.

REFERENCES

[1] D. S. Slepian and J. Wolf, "Noiseless coding of correlated information sources", *IEEE Trans on Information Theory*, Vol.19, No.4, 1973 July, pp.471-480.

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

[3] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed video coding", *Proceedings of the IEEE*, Vol.93, No.1, 2005 January, pp.71-83.

[4] A. Aaron, S. Rane, Zhang Rui, and B. Girod, "Wyner-Ziv coding for video: applications to compression and error resilience", *Proceedings of Data Compression Conference (DCC)*, 2003 March, pp.93-102.
[5] R. Puri, A. Majumdar and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder", *IEEE Trans on Image Processing*, 2007 October, Vol.16, No.10, pp.2436-2448.
[6] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation", *Proceedings of Picture Coding Symposium (PCS)*, 2007.
[7] F. Pereira, "Distributed video coding: Basics, main solutions and trends", *IEEE International Conference on Multimedia and Expo (ICME)*, 2009 Junuary, pp.1592-1595.
[8] T. Wiegand, G.J. Sullivan, G. Bjontegaard, A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Trans on Circuits and Systems for Video Technology*, 2003 July, Vol.13, No.7, pp.560-576.
[9] S. Klomp, Y. Vatis, and J. Ostermann, "Side information interpolation with sub-pel motion compensation for Wyner-Ziv decoder", *Proceedings of International Conference on Signal Processing and Multimedia Applications (SIGMAP)*, 2006 August, pp.520-528.
[10] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding", *Proceedings of European Signal Processing Conference (EURASIP 05)*, 2005 July.
[11] L. Alparone, M. Barni, F. Bartolini, and V. Cappellini, "Adaptively weighted motion vector median filters for motion fields smoothing", *Proceedings of IEEE International Conference*, Vol.4, 1996 May, pp.2267-2270.
[12] L. Yin, R. Yang, M. Gabbouj, Y. Neuvo, "Weighted median filters: a tutorial", *IEEE Trans on Circuits and Systems II: Analog and Digital Signal Processing*, Vol.43, No.3, 2004 March, pp.157-192.
[13] H. Ai-Mei, T.Q. Nguyen, "A multistage motion vector processing method for motion-compensated frame interpolation", *IEEE Trans on Image Processing*, Vol.17, No.5, 2008 May, pp.694-708.
[14] G. Dane, T.Q. Nguyen, "Motion vector processing for frame rate up conversion", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP '04)*, Vol.3. 2004 May, pp.309.

Jun Guo

He received his M.S. degrees in Electrical and Computer Engineering from Illinois Institute of Technology in 2010. And now he is undertaking a doctorate course as a member of the multimedia communications lab at Illinois Institute of Technology. His research interests include Video Processing, Multimedia Communication and Anomaly Behavior Analysis in Video Surveillance.

Joohee Kim

Joohee Kim received her B.S. and M.S. degrees in Electrical Engineering from Yonsei University, Seoul, Korea in 1991 and 1993, respectively. She received her Ph.D. degree from the Georgia Institute of Technology, Atlanta, GA, in 2003. From 1993 to 1997, she was with Korea Telecom Research Laboratories as a Research Engineer. She joined the Samsung Advanced Institute of Technology in 2003 as a Senior Research Engineer and developed various video coding algorithms. From 2005 to 2008, she was an Assistant Professor in the Department of Information and Communication Engineering at Inha University in South Korea. She is currently an Assistant Professor in the Department of Electrical Engineering at the Illinois Institute of Technology. Her current research interests include image and video signal processing, multimedia communications, wireless video, 3D video coding, and video analysis.