View-Specific Based Error Concealment Scheme for Multi-view Plus Depth Video

Ran Ma[1,2(✉)], Xiangyu Hu[1,2], Deyang Liu[1,2], Yu Hou[1,2], and Ping An[1,2]

[1] School of Communication and Information Engineering,
Shanghai University, Shanghai 200072, China
`maran@shu.edu.cn`
[2] Key Laboratory of Advanced Display and System Application,
Ministry of Education, Shanghai 200072, China

Abstract. Multi-view video plus depth (MVD) is an efficient format of 3D video. MVD video can be encoded by either H.264/AVC or HEVC standard to gain higher compression ratio, which benefits their broadcasting over the Internet. However, the encoded and transferred MVD video tends to develop worse visual quality degradation caused by lossy network channel. Therefore, error concealment is here to help refrain this problem. In this paper, we propose a classification-related error concealment method for MVD video. Within our method, motion-based classification is used to judge whether the corrupted blocks are static or not. If so, the blocks from reference frames are adopted to conceal the corrupted blocks. Otherwise, view-specific based concealment procedures, which are designed in accordance with view features, are used to conceal the corrupted blocks in different views. Experimental results on AVC-based Test Model (ATM) show the superiority of this concealment scheme to several other error concealment methods in PSNR along with acceptable execution time increase.

Keywords: Multi-view video plus depth · AVC-based Test Model · Error concealment · View-Specific

1 Introduction

Customers' passion for 3D video has constantly driven the advancement of 3D video coding technology. Two major 3D video formats have emerged for this demand, which are multi-view coding (MVC) and multi-view plus depth (MVD) [1]. The MVD format wins the other one at reconstructed visual quality of complex scene due to its extra depth maps that help synthesize virtual views, thus gaining increasing focus of the academic research and commercial solutions. However, compared to the traditional 2D video, the transferred MVD video is prone to develop heavier quality degradation caused by lossy networks for its extra views in single frame. Therefore, error concealment (EC) on received MVD videos is rather necessary to help mitigate this problem.

Related works about error concealment on MVD date back to early 2007. Depth-based boundary matching algorithm (DBMA) [2] was proposed, in which boundary matching algorithm (BMA) [3] was used to find candidate motion vector (MV) in depth

view for recovering the corrupted texture view block. Similar works appeared lately, where MVs of the corrupted blocks in texture/depth view were fixed by using their correspondence in depth/texture view [4, 5]. However, the inconsistency between MVs in texture view and their correspondence in depth view was ignored in the above three methods. To overcome this, certain kinds of block classification were applied for better performance. Method in [6] classified the corrupted blocks into homogeneous blocks and boundary blocks, then used BMA and decoder motion vector estimation (DMVE) [7] to fix the homogeneous or boundary blocks respectively. It achieved better concealment quality compared to previous ones, but still lacked further exploiting relations between texture and depth views in MVD video. Methods in [8, 9] predicted the corrupted blocks' prediction models by exploring relations between the 2 views, then proposed different concealment method for different prediction models. Furtherly, they gained increased concealment quality, but with the unacceptable computation time addition.

Our proposed method, which was enlightened by [9], makes further improvements on concealment efficiency. Instead of dividing corrupted blocks into simple or complex ones as described in [9], a second motion-based block type classification is carried out in our purposed view-specific methods. This classification varies in each of the views, catering each view' own characteristics. In our method, firstly, corrupted blocks are classified into two major kinds, static blocks and motional blocks. Static blocks are concealed by simply reusing collocated contents (MVs or pixel patches) from their reference frames, which largely reduce the complexity of our method. The concealment of motional blocks is related to the unique characteristics of the views they belong to. Specifically, motional blocks in different views should be concealed in view-specific ways for better overall quality enhancement and lower computation time. As texture views include complex color pattern and splines, for blocks in texture views, further size classification is executed instead to exploit the spatial relations of each block for better concealment quality. Contrary, depth views only contain simple objects and splines, so blocks in depth views should receive a secondary MV-related classification to further exploit temporal relations inside the video, attempting to reduce the overall time cost of our proposed method.

This paper is organized as follows. Section 2 describes the proposed hybrid error concealment method for MVD video. Experimental results are presented and analyzed in Sect. 3. Section 4 is the final conclusion part.

2 The Proposed Error Concealment Method

2.1 Frame Structure of MVD

Figure 1 describes the overall frame structure of typical multi-view plus depth (MVD) video, in which 6 views are included in a single video frame. The texture center (TC) view can be regarded as single frame in 2D videos. The depth center (DC) view that accompanies the TC view can increase the synthesized view quality during the display process. The texture/depth left (TL/DL) views and the texture/depth right (TR/DR) views provide more perspective angles of the MVD video, giving us more vivid 3D experience. These views are

processed with various order by different codecs. For example, in H.264/AVC codecs, TC view is encoded first, then DC view, followed by DL and DR views and finally TL and TR views. On the HEVC codecs, TL and TR views are just ahead of DL and DR views.
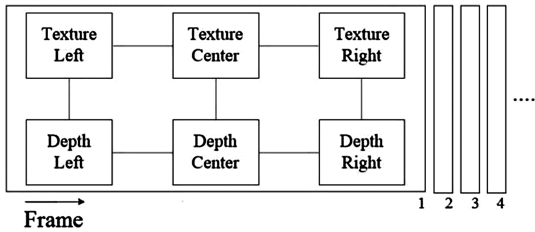

Fig. 1. MVD frame structure

Generally, like traditional 2D video frames, the texture views are color pictures that consist of complicated color patterns, multiple object edges and complex texture splines, while depth views are grey-level pictures only contain large object contours and plain areas. Figure 2 is an example of these 2 views.


Fig. 2. Example texture and depth views of a MVD frame

2.2   The Initial Block Classification

For error concealment, the computational complexity is a crucial requirement for real-life application. Therefore, initial block classification is to help reduce complexity by judging what kind of blocks can be concealed with lower computational method, and this classification of the corrupted blocks is firstly implied in every view.

As is shown in Fig. 3, corrupted blocks are classified into static and motional blocks. Static blocks are detected in large uniform regions with less details. Concealing these blocks with simpler method won't cause apparent video quality loss. Meanwhile, the time cost for the whole method can be largely decreased due to this. Contrarily, the motional blocks are discovered in regions with complex splines and figures, and should be concealed with the more effective, but complicated methods to ensure the overall concealment quality. In our case, the motional blocks are processed by the proposed view-specific methods for better quality assurance.

Corrupted Block

Static   Classification   Motional

Copy from
Reference
Frame

View-Specific
Methods

End
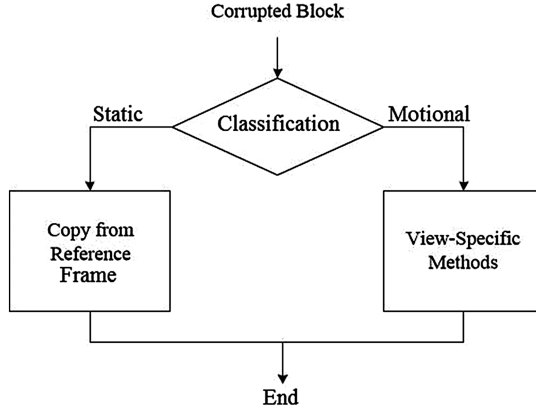
Fig. 3. General structure of proposed method

The initial classification process is as follows. Let the corrupted block be MB_0 and its eight-neighboring blocks be MB_1 to MB_8. The average of all the correctly received motion vectors (mv_i) from MB_1 to MB_8 is calculated according to Eq. 1. Where, the parameter p_i is used to mark whether the MB_i is correctly received or not (equals 1 if the MB_i is received correctly, otherwise it equals 0).

$$mv_{avg} = \frac{\sum_{i=0}^{8} mv_i \times p_i}{\sum_{i=0}^{8} p_i} \tag{1}$$

If this average vector (mv_{arg}) is 0, this corrupted block is a static block and is directly recovered by using its collocated block in the reference frame or view. Otherwise, it should be the motional bock.

Motional block usually indicates the movements of foreground objects and has its unique features inherited from the view who owns it. These features, in fact, represent the differences as well as similarities in each view according to the previous work by Liu [8, 9]. One of features for similarities is that motion vectors in TC or DC view should be identical to those in TL/TR or DL/DR views. On the other hand, the features for splines and textures can be greatly differed in different views, as we can find those to be much complex in texture views but very simple, even not exist in depth views.

Due to these features, view-specific concealment methods for motional blocks are proposed for better complying with view integrity, so as to acquire better concealment quality. Moreover, for the purpose of better trade-off between complexity and quality, at the beginning of the methods, a second block classification is carried out. TC, DC, DL/DR and TL/TR views are assigned to their respective view-specific methods described as follows.

2.3   Texture Center (TC) View-Specific Method

TC view usually has the biggest effect on the overall recovered video quality, so the efficient and powerful concealment should be applied to ensure best quality. For most codec standards, the encoding process of TC view is independent with other views in MVD video. Additionally, the large uniform regions are mostly presented by comparably big sized blocks (or SKIP/merge block), and those regions with many details are likely segmented by small sized ones. Thus, the procedure for concealment motional in TC view is as follows.

What comes first is the second classification for motional blocks in TC view. Motional block has to be judged as big or small size, determined by the size of its neighboring blocks. In case of H.264/AVC, SKIP, 16×16, 16×8 and 8×16 blocks belong to big size blocks because they probably present single spline and color. And 8×8, 4×8, 8×4 & 4×4 blocks are small size blocks because complex textures may be presented in these blocks.

Big size blocks usually have perfect resemblance in the reference views according to [9], so for its concealment, OBMA method is applied on previous and next reference views to get two candidate MVs. Then the best one is chosen by lowest minimum square sum of difference (SSD), to finally conceal the block. SSD determines the best vector that has the minimum difference between the outer bound of corrupted block and the outer bound of the regions it refers to.

Small size blocks usually contain more splines and colors, so it's necessary to segment these blocks into 4×4 sub-blocks and recover them separately. The vector mv_{arg} by Eq. 1 and the candidate vector by DMVE [7], which is applied on present view, are prepared, and the optimal one is chosen using SSD to finish concealing each sub-blocks.

2.4   Depth Center (DC) View-Specific Method

Since both DC and TC views present the same scene, many regions in DC view have strong motion homogeneousness with their collocated ones in TC View. This makes the reduction of method's complexity possible while maintaining its overall recovering quality. During the DC view-specific method, motional block's vector mv_{arg} by Eq. 1 and the one from its collocated block in TC view are subtracted for the second block classification in DC view method.

If the subtraction is 0, it is classified as consistent block, which means the motion similarity between the block and its collocated one in TC view is very likely to valid. Using contents referred by mv_{arg} to conceal this block is effective enough. Otherwise, it should be classified as inconsistent one, and both the spatial and temporal correlation of this block should be exploited. In most codec systems, two temporal reference views and one spatial reference view are referenced when processing DC view, which is described in Fig. 4.
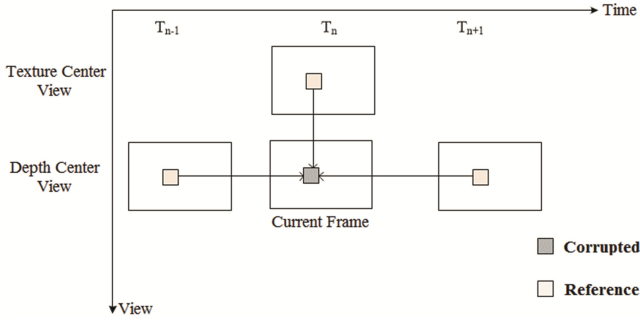
Fig. 4. Reference map of DC view

For dealing with blocks whose subtraction are not 0, three candidate vectors (mv_{argTC}, mv_{argDC+}, mv_{argDC-}) are calculated by using Eq. 1 on its collocated blocks in TC view at current frame, as well as in DC views at post and next frame, respectively. Then the optimal of these three vectors is chosen by SSD denoted in Sect. 2.3. Finally, the block is recovered using OBMA method on the view where the optimal vector is from.

2.5 Depth Left/Right (DL/DR) View-Specific Method

In most codec standard, DL/DR view has three reference views: DL/DR view at post and next frame, and DC view at current frame. Moreover, similar to DC view, DL/DR view has only contours of scene objects and large plain areas. This indicates strong relations between the motional block and its neighbors. Thus, simple concealment may achieve the satisfying recovering quality, as well as largely-reduced complexity.

DL/DR view-specific method is described as follows. For all motional blocks in DL/DR view, first, three average vectors, calculated by Eq. 1 on its collocated blocks in DC view at current frame, and in DL/DR views at post and next frame each, are prepared. Then, the optimal vector is chosen by lowest SSD from these three average vectors. Finally, the block is concealed using regions where the optimal vector refers to.

2.6 Texture Left/Right (TL/TR) View-Specific Method

Like DL/DR view, TL/TR view also has three reference views: DL/DR view at current frame and TL/TR view at post and next frame, which is similar to DC view. Furthermore, TL/TR view contains texture patterns while DL/DR view doesn't. Therefore, compared to DL/DR view, the concealment of TL/TR view need some additional steps to receive plausible quality.

Firstly, the block classification for TL/TR view method relies on motion subtraction. The subtraction of the average vectors, one acquired by Eq. 1 from the block itself and the other from its collocated block in DL/DR view, is calculated. If this subtraction is 0, the corrupted block is concealed with regions this vector refers to. Otherwise, OBMA is used to find the two best candidate vectors from TL/TR views at post and next

reference frame each, then the weighted average of these two is calculated, and finally, the optimal one is chosen between this weighted average and the MV of the collocated block in DL/DR view by lowest SSD. Once the optimal MV is obtained, the corrupted block can be recovered using pixel regions it refers to.

3   Experiments

Our experiment is conducted on AVC-based Test Model (ATM). Three MVD video sequences are used in the experiment: *Street*, *GT_Fly* and *Dancer*. All of these sequences have the resolution of 1920×1088. Quantization parameters in both texture and depth view are configured as 28. The Hierarchy-B prediction structure and P-I-P view prediction structure are separately applied at intra-view prediction and inter-view prediction. I frame is assumed to be correctly received and the Flexible Macroblock Ordering (FMO) [12] is enable at NAL level.

To validate the improvements of the proposed error concealment scheme, its performance is compared with the traditional OBMA [11] method's and Liu's [9] method's, which is also implemented on ATM.

Table 1 shows PSNR results which represent the overall concealment quality for the above 3 methods. Each of the test videos had suffered from 3%, 5% and 10% packet loss rate (PLR) and was recovered by OBMA, Liu's and the proposed method respectively. Also, the averages for each methods' performance at each view is also listed. Our results prove that the proposed scheme outperforms the other two schemes for better concealment quality in all views. Compared to OBMA, the proposed scheme has PSNR increase in average 0.24 dB, 0.31 dB and 0.38 dB for 3%, 5% and 10% PLR, respectively. For Liu's method, the increase is in average 0.12 dB, 0.22 dB and 0.23 dB for 3%, 5% and 10% PLR. In case of 3% PLR, the TL/TR view-specific method achieves the biggest PSNR average increases, which are 0.3 dB to OBMA, and 0.2 dB to Liu's method. Meanwhile, for 5% and 10% PLR, DC view-specific method instead, gains biggest increases, which are 0.21 dB and 0.35 dB to Liu's method at 5% and 10% PLR respectively. These results indicate better designed view-specific methods for TL/TR view and DC view, compared to ones for the other views. Since both methods carry out the second block classification using motion subtraction, one can infer that the classification by motion subtraction is more capable of eliminating scene inconsistency that exists in one view.

Another interesting fact from the results is that, the PSNR for proposed method has the lowest decrease with the PLR increases. From 3% to 10% PLR, the average PSNR decrease for OBMA and Liu's method are 2.08 dB and 2.04 dB respectively, while 1.94 dB is decreased for the proposed method. So, the proposed method has better error resilient performance compared to OBMA and Liu's.

Table 1. PSNR comparison under different packet loss rate

Sequence & view		PLR								
		3%			5%			10%		
		OBMA (dB)	Liu (dB)	Proposed (dB)	OBMA (dB)	Liu (dB)	Proposed (dB)	OBMA (dB)	Liu (dB)	Proposed (dB)
Street	TC	36.56	36.83	36.93	35.73	36.08	36.12	34.14	34.62	34.83
	TL	34.13	34.21	34.46	33.03	33.15	33.41	30.75	30.83	30.92
	TR	34.11	34.20	34.33	33.28	33.38	33.50	30.65	30.70	30.85
	DC	40.12	40.42	40.54	39.74	39.76	40.42	38.67	38.70	39.15
	DL	39.44	39.47	39.57	39.11	39.19	39.29	37.54	37.60	37.71
	DR	39.89	40.04	40.13	39.20	39.32	39.40	37.53	37.90	38.15
GT_Fly	TC	37.99	38.06	38.17	37.29	37.44	37.59	35.87	36.07	36.22
	TL	36.91	36.95	37.19	36.26	36.32	36.62	34.64	34.68	35.14
	TR	37.16	37.18	37.45	36.30	36.36	36.77	34.67	34.71	35.29
	DC	41.45	41.49	41.64	41.33	41.42	41.60	40.88	40.97	41.23
	DL	41.79	41.82	41.86	41.48	41.56	41.70	40.90	40.96	41.17
	DR	41.91	41.90	41.98	41.55	41.57	41.65	40.89	40.91	41.08
Dancer	TC	34.10	34.21	34.35	33.19	33.24	33.54	31.00	31.22	31.52
	TL	32.04	32.14	32.32	31.12	31.23	31.31	29.63	29.76	29.91
	TR	31.93	32.18	32.30	31.21	31.26	31.33	29.71	29.91	30.01
	DC	38.92	39.06	39.17	38.83	38.68	39.03	37.83	37.92	38.04
	DL	37.29	37.35	37.49	36.06	36.24	36.47	34.90	35.09	35.32
	DR	37.22	37.33	37.37	35.79	35.88	36.16	35.24	35.58	35.76
Average	TC	36.22	36.37	36.48	35.40	35.59	35.75	33.67	33.97	34.19
	TL	34.36	34.43	34.66	33.47	33.57	33.78	31.67	31.76	31.99
	TR	34.40	34.52	34.69	33.60	33.67	33.87	31.68	31.77	32.05
	DC	40.16	40.32	40.45	39.97	39.95	40.35	39.13	39.20	39.47
	DL	39.51	39.55	39.64	38.88	39.00	39.15	37.78	37.88	38.07
	DR	39.67	39.76	39.83	38.85	38.92	39.07	37.89	38.13	38.33

Table 2 shows us the OBMA, Liu's & proposed methods' execution time, along with the Liu's & proposed methods' time increments compared to classic OBMA method. Apparently, the time consumption for our proposed method is just slightly higher than Liu's & OBMA method with acceptable, even negligible time increments, despite the classification and view-specific method applied on each view, making the proposed method be more complicated than the other two.

Table 2. Comparison for time consumption under different packet loss rate (PLR)

Sequence	PLR (%)	OBMA	Liu		Proposed	
		Time (s)	Time (s)	Increment (s)	Time (s)	Increment (s)
Street	3	4.63	4.80	0.17	4.78	0.15
	5	4.75	5.04	0.29	4.83	0.08
	10	4.83	5.13	0.30	5.01	0.18
GT_Fly	3	4.63	4.78	0.15	4.81	0.18
	5	4.75	4.83	0.08	4.89	0.12
	10	4.83	5.02	0.19	5.10	0.27
Dancer	3	4.99	5.21	0.23	5.22	0.24
	5	5.23	5.41	0.18	5.48	0.25
	10	5.40	5.99	0.59	5.93	0.53

This is mainly due to the classification of the motional or static blocks applied before view-specific methods. The classification gives a chance to most of these corrupted blocks of much easier copying concealment, if they are classified as static. The experiment results give the proof of the certain extent of computation reduction that the classification brings to proposed method.

4   Conclusion

We proposed a new view-specific based error concealment scheme for MVD. During the overall method, two unique classifications are conducted in turn, for judging the most suitable concealment for each corrupted block. Also, different view-specific methods are executed in different views according to different view features. The first classification helps us reducing time that is wasted on concealing corrupted blocks. The second classification helps improving the video quality for better catering each view's own characteristics. Compared to previous methods, we achieved a better trade-off between concealment quality and time consumption, which can be proved by our experiment results.

References

1. Merkle, P., Smolic, A., Muller, K., Wiegand, T.: Multi-view video plus depth representation and coding. In: 2007 IEEE International Conference on Image Processing, vol. 1, pp. 201–204 (2007)
2. Yan, B.: A novel H.264 based motion vector recovery method for 3D video transmission. IEEE Trans. Consum. Electron. **53**(4), 1546–1552 (2007)
3. Lam, W.M., Reibman, A.R., Liu, B.: Recovery of lost or erroneously received motion vectors. In: 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 417–420 (1993)
4. Hewage, C.T.E.R., Worrall, S.T., Dogan, S., Kondoz, A.M.: A novel frame concealment method for depth maps using corresponding color motion vectors. In: 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, pp. 149–152 (2008)
5. Hewage, C.T.E.R., Martini, M.G.: Joint error concealment method for backward compatible 3D video transmission. In: 2011 IEEE 73rd Vehicular Technology Conference (VTC Spring), pp. 1–5 (2011)
6. Liu, Y., Wang, J., Zhang, H.: Depth image-based temporal error concealment for 3-D video transmission. IEEE Trans. Circ. Syst. Video Technol. **20**(4), 600–604 (2010)
7. Zhang, J., Arnold, J.F., Frater, M.R.: A cell-loss concealment technique for MPEG-2 coded video. IEEE Trans. Circ. Syst. Video Technol. **10**(4), 659–665 (2000)
8. Liu, D., Ma, R., Shi, N.D., et al.: An error concealment method for multi-view video plus depth. J. Optoelectron. Laser **24**(8), 1548–1591 (2013)
9. Liu, D., Ma, R., Shi, N.D., et al.: An error concealment method for multi-view video plus depth coding. J. Optoelectron. Laser **25**(8), 756–763 (2014)

10. Su, W., Rusanovskyy, D., Hannuksela, M.M., Li, H.: Depth-based motion vector prediction in 3D video coding. In: Picture Coding Symposium (PCS), pp. 37–40 (2012)
11. Thaipanich, T., Wu, P.-H.,. Kuo, C.-C.J: Video error concealment with outer and inner boundary matching algorithms. In: Proceedings of SPIE, vol. 6696, p. 669607. The International Society for Optical Engineering (2007)
12. Dhondt, Y., Mys, S., De Zutter, S., Van de Walle, R.: An alternative scattered pattern for flexible macroblock ordering in H.264/AVC. In: Eighth International Workshop on Image Analysis for Multimedia Interactive Services, pp. 57–57 (2007)