DEPTH-ASSISTED ERROR CONCEALMENT FOR INTRA FRAME SLICES IN 3D VIDEO

Meng Yang[1], *Yuhong Yang*[2], *Pamela Cosman*[3]

[1]AIAR, Xi'an Jiaotong University, Xi'an, P.R. China
[2]Shanghai Jiaotong University, Shanghai, P.R. China
[3]ECE, University of California at San Diego, CA, USA

ABSTRACT

We propose a depth-assisted error concealment method for slice loss in intra frames of 2D+depth video sequence. Intra frames in the 2D view sequence are offset from intra frames in the depth sequence to guarantee the corresponding frame in the other sequence is not also intra mode. Then for a slice loss in an intra frame in the 2D view sequence, the motion information is extracted from the depth sequence to conceal the slice loss using boundary matching. Experimental results show that the proposed method provides improved performance over existing methods both for PSNR results and computational complexity at the decoder.

Index Terms— Error concealment, Slice loss, Intra frame, 3D video coding.

1. INTRODUCTION

2D+depth encoding of 3D video shows promise due to its low bit rate, compared with multi-view video (MVC) techniques. In 2D+depth 3D video, an additional depth sequence associated with the 2D video sequence is captured or computed, compressed, and then transmitted independently. At the decoder, depth-image-based rendering (DIBR) [1] can be used to reconstruct the 3D video based on the 2D video stream and the associated depth stream.

All video techniques suffer losses during transmission. The dominant solutions at the decoder include: spatial error concealment (SEC) which is mainly for intra mode (I frame), and temporal error concealment (TEC) which is mainly for inter mode (P, B frame). One of the well-known TEC methods is the boundary matching algorithm (BMA) [2]. It estimates several candidate motion vectors (MVs) of the lost region from the MVs of the surrounding regions, and then finds the best region from the previous frames. BMA requires the MV information of the current frame, which is only suitable for inter frames. As motion estimation is not implemented in intra frames, the dominant solution is SEC, such as spatial interpolation [3] [4]. However, most SEC methods perform less well than TEC, especially for consecutive MB losses in natural video sequences. For intra mode, a better solution is to utilize the temporal correlation of the sequence. One method is to implement motion estimation at the decoder for the current intra frame using the decoded inter frames. However, the computational complexity is unacceptable for decoding, even though it can be optimized as with the decoder motion vector estimation (DMVE) [5]. Another efficient method is data hiding [6], which executes motion estimation for consecutive intra frames at the encoder and then imperceptibly embeds MV into the coefficients of other MBs in the same intra frame. If MB loss occurs in the intra frame at the decoder, the MV of this MB may be extracted from the current intra frame. However, this method relies on both the decoder and the encoder, and it requires extra computational complexity at the encoder and extra bits.

The loss problem can be reconsidered in 3D video transmission, because there is a close relation between the 2D sequence and depth sequence, especially for the motion information [7]. Yan [8] and Liu [9] utilize the depth information to further optimize the conventional BMA method for the packet loss problem in inter frames with good results. However, the method does not work for intra mode, because there is no MV information in either the 2D sequence or the depth sequence for the intra frame.

In this paper, we consider slice loss for intra frames in 3D video transmission, where one slice (packet) is one MB row. We try to extract motion information from the depth sequence, and conceal the lost packets of the intra frames in the 2D sequence by TEC without any increase of computational complexity.

The rest of this paper is organized as follows. In Section 2, we review the BMA method. In Section 3, the proposed error concealment method for intra frames in 3D video sequences is introduced. In Section 4, we evaluate the proposed method, and we conclude the paper in Section 5.

2. BOUNDARY MATCHING ALGORITHM

The BMA method, which is used for inter frames, attempts to recover the MV of each lost MB (denoted MB_{2D}^{lost}) of a video sequence using neighbor MBs (denoted MB_{2D}^{i}, i=0,1, 2,…), as shown in Fig. 1. Let the size of an MB be $N \times N$.

Let MV_{2D}^{i} be the candidate MVs which correspond to MB_{2D}^{i} (i=0,1,2,…). The optimal MV is selected by minimizing the side match distortion between the internal and external boundaries of the reconstructed MBs. Let the pixel values of last reconstructed frame n-1 and current frame n be denoted $P(x,y,n-1)$ and $P(x,y,n)$, where (x,y) denotes the spatial coordinate. The side match distortion [2] is computed as:

$$\begin{aligned} SSD_{2D} = &\sum_{x=x_0}^{x_0+N-1}\Big[\big(P(x,y_0-1,n)-P(x+v_x,y_0+v_y,n-1)\big)^2 \\ &+\big(P(x,y_0+N,n)-P(x+v_x,y_0+N-1+v_y,n-1)\big)^2\Big] \\ &+\sum_{y=y_0}^{y_0+N-1}\Big[\big(P(x_0-1,y,n)-P(x_0+v_x,y+v_y,n-1)\big)^2 \\ &+\big(P(x_0+N,y,n)-P(x_0+N-1+v_x,y+v_y,n-1)\big)^2\Big]. \end{aligned}$$

where (x_0,y_0) denotes the coordinate of the top-left pixel in the damaged MB_{2D}^{lost}, and (v_x,v_y) is the MV candidate. If the MV information was lost but the prediction residue information was not lost, the optimal replacing MB is summed with the prediction residue of MB_{2D}^{lost}. In case all information of MB_{2D}^{lost} is lost, the residue of a neighbor MB can be used instead, or the replacing MB can be used alone.

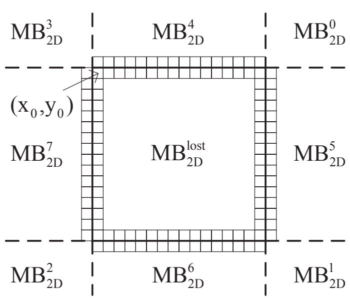


Fig. 1 Boundary match criterion for MV selection.

3. PROPOSED ERROR CONCEALMENT FOR INTRA FRAME SLICES

In order to conceal packet errors in the 2D sequence using the motion information in the depth sequence, we need similarity of motion information between the 2D sequence and the depth sequence, and availability of the MVs of the corresponding frame in the depth sequence.

In a 2D+depth sequence, the depth sequence is extracted from the 3D scene to assist with the 2D sequence. For most natural objects in a 3D scene, the 2D component moves along with the depth component. So the relationship between 2D and depth motion information can be exploited efficiently. The exception scenario is when a 2D object (no depth difference for all pixels) moves like 2D video without any depth changes. Denoting the MVs of an MB in frame n of the 2D sequence and of the depth sequence as $MV_{2D}=(v_x^{2D},v_y^{2D})$ and $MV_{depth}=(v_x^{depth},v_y^{depth})$ respectively, we define the MV similarity $MVSIM_n$ $(0\le MVSIM_n\le 1)$ for a frame in the two sequences as:

$$MVSIM_n=1-\frac{1}{T}\sum_{\substack{MV_{2D}\in\Delta_n^{2D}\\ MV_{depth}\in\Delta_n^{depth}}}\left[\frac{\left(v_x^{2D}-v_x^{depth}\right)^2+\left(v_y^{2D}-v_y^{depth}\right)^2}{\left(v_x^{2D}\right)^2+\left(v_y^{2D}\right)^2}\right]$$

where Δ_n^{2D} and Δ_n^{depth} denote the MV sets of frame n in the 2D sequence and the depth sequence respectively. T denotes the number of MVs in a whole frame. The larger the value is, the more similar the MVs between the two sequences are. $MVSIM_n=1$ means the MVs of the two sequences are the same. The average MV similarity value of all frame pairs in a 2D sequence and corresponding depth sequence are shown in TABLE I to verify the similarity between their MV information.

TABLE I. MV Similarity between 2D and depth sequences

Sequence	Newspaper	Bookarrival	Cafe
MVSIM	0.7893	0.7368	0.8732
Sequence	Pantomime	Mobile	Balloons
MVSIM	0.7671	0.856	0.48

In this paper, we use the motion information from the depth sequence directly without exploring their similarity further. For packet loss in the 2D sequence, all the neighboring and co-located MVs of the corresponding depth frame are exploited directly as MV candidates.

To do this, first we need to extract MVs from the depth sequence. If the corresponding frame in the depth sequence is also intra mode, there is no MV information that can be used. We impose a temporal offset to the intra frames of the depth sequence, as shown in Fig. 2. We denote the GOP size as G and the GOP offset as G_{offset} ($0<G_{offset}<G$). An offset is imposed to the depth sequence starting from the 2nd GOP. That means the size of the 1st GOP in the depth sequence is $G\text{-}G_{offset}$, and the size of the other GOPs is G. The frame in the depth sequence that corresponds to the intra frame in the 2D sequence is inter coded.

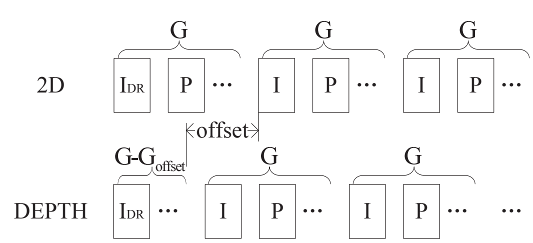
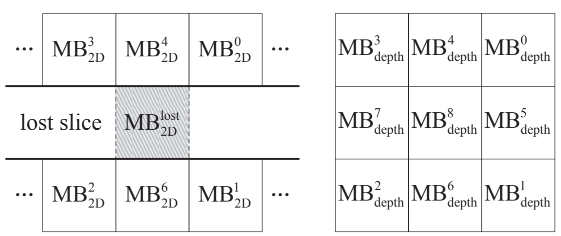


Fig.2 Organization of 2D and depth sequences.

In conventional error concealment for intra frames, only spatial information can be used. In Fig. 3(a) for example, there is a slice loss in frame *n* of the 2D sequence, which is intra mode. Each macroblock MB_{2D}^{lost} in the lost slice can be concealed by the 2 neighbor macroblocks MB_{2D}^{i} (i=4,6) or the 6 ones MB_{2D}^{i} (i=0,2,...,4,6). In the proposed method, both the spatial correlation in the intra frame of the 2D sequence and the temporal correlation in the depth sequence are combined. The co-located MBs in the depth sequence are denoted MB_{depth}^{i} (i=0,1,...,8), as shown in Fig. 3(b). Frame *n* in the depth sequence is inter coded. The MVs of MB_{depth}^{i} (i=0,1,...,8) are extracted and denoted MV_{depth}^{i} (i=0,1,...,8). If the slice in the depth sequence that corresponds to the lost slice in the 2D sequence is also lost, we only use 6 MVs as MV candidates, that is MV_{depth}^{i} (i=0,1,...,4,6). If the 2D sequence and the depth sequence are transmitted independently, the corresponding slice in depth sequence may not be lost, and then all 9 MVs of the depth sequence are used as MV candidates. In this paper, we only consider the latter scenario.



Frame *n* in 2D: Intra (a) — Frame *n* in DEPTH: Inter (b)

Fig. 3 Correlation between 2D frame and DEPTH frame, (a) Spatial MB candidates in 2D frame; (b) Spatial MB candidates in depth frame.

Then the steps of the proposed method are as follows.

1) Extract the MV_{depth}^{i} $(i = 0,1,...,8)$ from the depth sequence as the MV candidates.

2) Find the optimal MV from the MV candidates using BMA, and conceal the error macroblock MB_{2D}^{lost}.

4. EXPERIMENTS

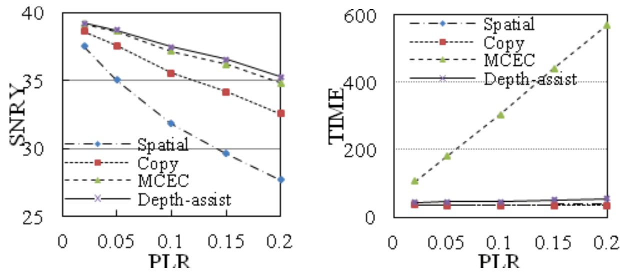
We implement the proposed scheme based on H.264/AVC codec reference software (JM16.2). The GOP size of the sequence is set as $G = 4$ to generate more intra frames in a sequence for experiment, and the format of the 2D sequence is simply set as "IPPPIPPP...". The offset parameter of the depth sequence is $G_{offset} = 1$. That means the format of the depth sequence is "IPPIPPPI...". The motion search range is 16. The slice is defined as an MB row. Six standard sequences are tested: Newspaper (1024×768), Bookarrival (1024×768), Mobile (720×540), Pantomime (1280×960), and Cafe (1920×1080). The packet losses of the intra frames are simulated by the rtp_loss tool in the JM16.2 software. Four methods are compared for the slice loss problem in intra frames. 1) Spatial interpolation: Recover each pixel in the lost slice by linear interpolation using the nearest pixels in the top and bottom neighbor MBs. It is denoted "Spatial". 2) Zero motion estimation: Use the co-located MBs of the previous decoded frame to replace the MBs of the lost slice simply. It is denoted "Copy". 3) Decoder motion estimation combined with BMA: For each MB of the lost slice in current frame *n*, implement motion estimation again at the decoder for the 6 neighbor MBs (Fig. 3(a)) as MV candidates, and then conceal the lost region by BMA. It is denoted "MCEC". 4) Proposed method: It is denoted "Depth-assist". We test the PSNR result of the Y component (denoted "SNRY") and the decoding time (denoted "TIME") of each method to illustrate their performance.

We first test the result of each intra slice independently, and take the average value of all slice loss cases in all intra frames of the input sequence. The first IDR frame of the sequence is ignored. The results are shown in Table II. Compared with spatial interpolation and zero motion estimation, the SNRY result of the proposed method is always much better, and the decoding times of these three methods are almost equivalent. Compared with decoder motion estimation, the SNRY result of the proposed method is a little better, and the decoding time of the proposed method is much less. The MVs of the depth sequence are not identical to those of the 2D sequence. However, the SNRY result of Depth-assist is always better than that of MCEC possibly because only 6 MV candidates can be used for MCEC whereas 9 ones are used for the proposed method.

TABLE II. Average SNRY and TIME comparison

Sequence		Spatial	Copy	MCEC	Depth-assist
Newspaper	SNRY	37.10	39.01	39.54	**39.59**
	TIME (ms)	60	60	290	88
Bookarrival	SNRY	37.83	38.45	38.78	**38.82**
	TIME (ms)	53	51	287	76
Mobile	SNRY	35.16	39.01	39.44	**39.47**
	TIME (ms)	29	29	176	42
Pantomime	SNRY	40.15	38.87	39.90	**40.23**
	TIME (ms)	120	119	600	173
Cafe	SNRY	40.29	41.47	41.90	**41.95**
	TIME (ms)	127	125	568	182

One important issue for video coding is the error propagation in the subsequent inter frames if the slice loss occurs in an intra frame. In order to verify the error propagation performance, we test packet loss rates (PLR) in all intra frames of 2%, 5%, 10%, 15%, and 20%. The average SNRY and TIME of the whole newspaper sequence are shown in Fig. 4. The SNRY result for PLR=10% is shown in Fig. 5 frame by frame (only the first 50 frames). These results are similar to Table II.



(a) SNRY (b) TIME

Fig. 4 Average result of "Newspaper" sequence.

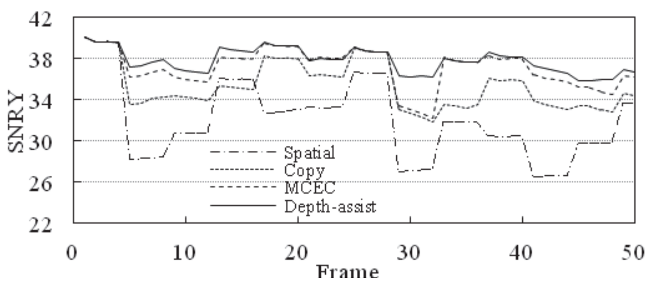


Fig. 5 SNRY result of "Newspaper" sequence.

Another advantage of the proposed method lies in the bit rate stability of the coded streams. The output bit rate of coded 2D+depth sequences is the sum of that of the 2D sequence and that of the depth sequence. For conventional coding, the total bit rates fluctuate sharply between the intra and inter frames, because the bit rates for intra frames are higher for the inter frames in both the 2 streams. The result is shown in Fig. 6 denoted "0 offset". In the proposed method, intra frames in the 2D sequence are offset from intra frames in the depth sequence. This results in shifting part of the bit rate in the peak point to that in the trough point. This benefits the stability of the output stream. The results are shown in Fig. 6, where "1 offset" and "2 offset" stand for $G_{offset}=1$ and $G_{offset}=2$ cases respectively.

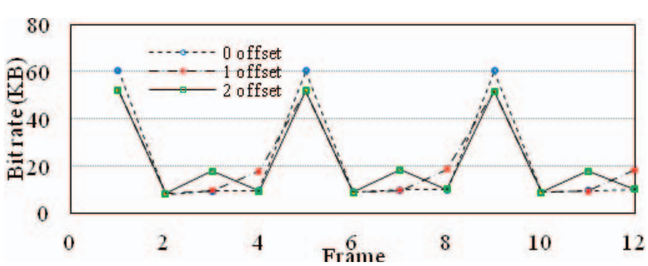


Fig. 6 Total bit rate of "Newspaper" 2D+depth sequence.

5. CONCLUSION

In this paper, we proposed a depth-assisted error concealment method for slice loss in intra frames of 2D+depth video. The correlation between the motion information of the 2D sequence and that of the depth sequence is exploited. Experimental results show that the proposed method has improved PSNR, reduced computational complexity at the decoder, and more stable output bit rate.

6. CONCLUSION

This work was supported in part by the Program 973 No. 2012CB316400, and NSFC No.61175010 of China.

7. REFERENCES

[1] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3DTV," in Proc. 11th Int. Soc. Opt. Eng. Conf. Stereoscopic Displays Virtual Reality Systems, San Jose, CA, USA, vol. 5291, pp. 93–104, Jan. 2004.

[2] W. Lam, A. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in Proc. ICASSP 1993, Minneapolis, MN, pp. 417–420, Apr. 1993.

[3] X. Li and M. Orchard, "Novel sequential error-concealment techniques using orientation adaptive interpolation," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 10, pp. 857–864, Oct. 2002.

[4] S. Hsia, "An edge-oriented spatial interpolation for consecutive block error concealment," IEEE Signal Processing Letters, vol. 11, no. 6, pp. 577-580, Jun. 2004.

[5] J. Zhang, J. Arnold, and M. Frater, "A cell-loss concealment technique for MPEG-2 coded video," IEEE Trans. Circuits Syst. Video Technol., vol. 10, no. 6, pp. 659–665, Jun. 2000.

[6] S. Chen and H. Leung, "A temporal approach for improving intra-frame concealment performance in H.264/AVC," IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 3, pp. 422–426, Mar. 2009.

[7] C. Hewage, S. Worrall, S. Dogan, and A. Kondoz "A novel frame concealment method for depth maps using corresponding color motion vectors," in Proc. 3DTV-CON'08, Istanbul, Turkey, pp. 149-152, May 2008.

[8] B. Yan, "A novel H.264 based motion vector recovery method for 3D video transmission," IEEE Transactions on Consumer Electronics, vol. 53, no. 4, pp. 1546-1552, Nov. 2007.

[9] Y. Liu, J. Wang, and H. Zhang , "Depth image-based temporal error concealment for 3-D video transmission," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 4, pp. 600–604, Apr. 2010.