

Optical Engineering

SPIDigitalLibrary.org/oe

Texture video-assisted motion vector predictor for depth map coding

Xiaoxian Liu
Yilin Chang
Zhibin Li
Junyan Huo

Texture video-assisted motion vector predictor for depth map coding

Xiaoxian Liu, Yilin Chang, Zhibin Li, and Junyan Huo
Xidian University, State Key Laboratory of ISN, Xi'an,
Shaanxi, 710071 China
E-mail: xiaoxian_liu@126.com

Abstract. A texture video-assisted motion vector predictor for depth map coding is proposed in this letter. Based on the analyses of motion similarity between texture videos and their corresponding depth maps, the proposed approach uses the motion vectors of texture videos and the median predictor jointly to determine the optimal predicted motion vector for depth map coding by employing a rate-distortion (R-D) criterion. Experimental results demonstrate that compared with the median predictor utilized in H.264/AVC, the proposed method can save the maximum and average bit rate as high as 4.89% and 3.68%, respectively, while guaranteeing the quality of synthesized virtual views. © 2011 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.3615644]

Subject terms: H.264/AVC; predicted motion vector; depth map coding.

Paper 101026LRR received Dec. 4, 2010; revised manuscript received Jun. 19, 2011; accepted for publication Jul. 7, 2011; published online Aug. 4, 2011.

1 Introduction

Three-dimensional video (3DV) has attracted people's attention as an emerging new media. Owing to limited transmission bandwidth, efficient compression of 3DV, which consists of texture video coding and depth map coding, is a key enabling factor for its applications. In Ref. 1, research on texture video compression has been carried out by adopting the interview correlations between different viewpoints besides temporal and spatial correlations. Also, many works have focused on depth map coding recently.^{2,3}

In Ref. 2, the motion estimation process was skipped by directly using the motion vectors (MVs) of texture videos, thus reducing the coding cost for depth maps. While the method in Ref. 3 generated three candidate modes and MVs for depth map coding using the motion information of texture videos. Inspired by these previous works, the motion information of texture videos was utilized in this letter for improving the depth map coding efficiency.

In H.264/AVC, to improve the video coding efficiency of interprediction, the median of the MVs of neighboring coded blocks was applied to predict the MV for the current coding block. Thus, the accuracy of predicted motion vector (PMV) highly affects the coding efficiency of motion-compensated prediction.⁴ Among the existing proposed methods for improving the efficiency of the predictive coding of MV, the competition-based PMV selection method proposed by Laroche⁵ has been adopted to the Key Technol-

ogy Area (KTA) due to its significant bit rate reduction for texture videos. Motivated by the MV-competition method,⁵ this letter proposed a texture video-assisted motion vector predictor for depth map coding.

2 Motion Correlation Analyses Between Texture Videos and Depth Maps

2.1 Motion Similarity Analysis between Texture Videos and Depth Maps

For the object belonging to the same macroblock, the MVs of texture videos and depth maps should be the same in a sense of physics. However, the texture video MVs cannot represent the physical displacement values of the object due to texture complexity. Thus, the MVs between texture videos and depth maps may be coincident when reflecting the displacement of the same object. While in other cases, since the pixels in the coding block belong to different objects, or even if they are located in the same object, due to the texture complexity, the MVs between texture videos and depth maps are not the same, even not similar, as shown in the following experiment.

In the experiment, texture videos and depth maps were encoded with H.264/AVC using four quantization parameters (QPs), i.e., 27, 32, 37, 42. The MVs of texture videos and depth maps for each 4×4 block were extracted and analyzed. An MV similar evaluation criterion is introduced in Eq. (1)

$$E_{mv}(i, j, k) = \begin{cases} 1, & \|\mathbf{MV}_t(i, j, k) - \mathbf{MV}_d(i, j, k)\| < T \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

where $\mathbf{MV}_t(i, j, k)$ and $\mathbf{MV}_d(i, j, k)$ are the MVs for block (i, j) in the k 'th picture of the texture videos and depth maps, respectively, and T is a predefined threshold and is set to 2. It is noted that the threshold T is only applied in motion similarity analysis and will not be used in the following experiments. When $\|\mathbf{MV}_t(i, j, k) - \mathbf{MV}_d(i, j, k)\| < T$, E_{mv} is set to 1, namely, MVs of the two blocks are similar; otherwise setting E_{mv} to 0, which means there may be quite different MVs of the two blocks. Then, the percentage of similar MVs of corresponding blocks between texture videos and depth maps was computed via Eq. (2)

$$P_{td} = \frac{16 \times \sum_{k=1}^{N_f} \sum_{i=1}^{H/4} \sum_{j=1}^{W/4} E_{mv}(i, j, k)}{H \times W \times N_f}, \quad (2)$$

where N_f represents the total number of pictures and $H \times W$ represents the image size. All the P_{td} values are shown in Table 1.

As seen from Table 1, the value of P_{td} becomes larger with the increase of QP value for a certain sequence. It is because there are much more skipped macroblocks (MBs) and 16×16 modes in depth map coding than texture video coding,³ while for texture video coding, the number of such MBs and modes increases for higher QP, thus resulting in the increased percentage of similar MVs between texture videos and depth maps. To summarize, when coding depth maps, the MVs of several blocks in texture videos can be used as PMVs for associated blocks in depth maps; while for other blocks with relatively large MV differences, the texture video MVs should not be used.

Table 1 Percentage of similar MVs between texture videos and depth maps.

Sequences	P_{td} (%)			
	QP = 27	QP = 32	QP = 37	QP = 42
Breakdancers view 02	59.02	67.80	76.36	84.60
Breakdancers view 04	59.42	68.91	77.39	85.71
Ballet view 02	75.64	80.84	86.70	90.12
Ballet view 04	77.01	82.42	87.40	89.81
Bookarrival view 08	77.47	81.39	83.89	86.35
Bookarrival view 10	75.17	79.39	83.48	86.51

2.2 Accuracy Analysis of Employing Texture Video MV as PMV for Depth Map

We compared the accuracy of using the median PMV and texture video MV as PMV for depth map, and the results are described in Table 2.

In the experiment, the MV for the depth map block is denoted as MV_d , and the median PMV and PMV based on texture video MV are PMV_m and PMV_t , respectively. $\|MVD_m\|$ represents the vector difference between MV_d and PMV_m , and $\|MVD_t\|$ represents the vector difference between MV_d and PMV_t . In Table 2, P_{tm} indicates the percentage of the blocks with $\|MVD_t\| - \|MVD_m\| \leq 0$ in all the blocks of the picture. From Table 2, the value of P_{tm} increases with the increase of QP value for the certain sequence, which is mainly attributable to the occurrence of more skipped MBs and 16×16 modes in texture video coding for higher QPs. Thus, we can draw the conclusion that for some blocks of depth maps, it is more accurate to select the MV of corresponding texture video as PMV for depth map than the median PMV.

3 Texture Video-Assisted Motion Vector Predictor for Depth Map Coding

Motivated by the analyses in Sec. 2, a texture video-assisted motion vector predictor for depth map coding was proposed.

Table 2 Comparison results of motion vector difference.

Sequences	P_{tm} (%)			
	QP = 27	QP = 32	QP = 37	QP = 42
Breakdancers view 02	15.55	25.93	34.73	46.19
Breakdancers view 04	18.93	28.68	42.26	53.50
Ballet view 02	62.92	75.67	86.89	92.35
Ballet view 04	63.28	79.51	87.86	92.63
Bookarrival view 08	77.82	84.51	87.98	92.10
Bookarrival view 10	72.03	77.83	85.05	90.54

In the proposed algorithm, both the median PMV and texture video MVs are adopted to predict MVs for depth maps.

In order to obtain the optimal PMV between the median PMV and the proposed PMV for depth maps, the rate-distortion optimization criterion⁵ is employed to select the PMV with the minimum rate-distortion (R-D) cost. The Lagrangian cost function is given by

$$J(\mathbf{S}, \mathbf{I}|\lambda) = D(\mathbf{S}, \mathbf{I}) + \lambda R(\mathbf{S}, \mathbf{I}), \quad (3)$$

where \mathbf{S} is the depth map data, \mathbf{I} is the coding parameter set including coding mode, motion information, etc., λ indicates the Lagrangian multiplier, $D(\mathbf{S}, \mathbf{I})$ and $R(\mathbf{S}, \mathbf{I})$ are the total distortion and rate, respectively, resulting from the coding of \mathbf{S} with a particular combination of coding options \mathbf{I} . It should be noted that the selection of PMV is included in \mathbf{I} . The optimal PMV can be obtained via

$$\mathbf{PMV}^* = \arg \min \{J[\mathbf{S}, \mathbf{I}(\mathbf{PMV}_i |_{i \in \{t,m\}})|\lambda]\}. \quad (4)$$

Similar with the MV-competition method in Ref. 5, the selected optimal PMV can give the minimum R-D cost for skip mode; while for other modes, the selected optimal PMV can guarantee the minimum motion vector difference.

4 Experimental Results and Analyses

Experiments were carried out using three 3DV sequences as shown in Table 3. In Table 3, the left view and right view were used to synthesize the virtual view in the position of the central view. Experiments were performed on the KTA test platform version 2.6r1 (KTA2.6r1). Context-based adaptive binary Arithmetic coding and four QPs (27, 32, 37, 42) were used.

The performance comparison among the proposed method, H.264/AVC, and the MV-competition (temporal PMV + median PMV) method⁵ is tabulated in Table 4. In Table 4, the Bjontegaard delta (BD) bit rate⁶ denotes the percentage of average bit rate savings at the same coding quality; while the BD peak signal-to-noise ration (PSNR)⁶ indicates average increase of PSNR at the same coding bit rate. As seen from Table 4, the achieved average BD bit rate is -3.68% compared with H.264/AVC. The percentages of selecting texture video MV as optimal PMV among skip and 16×16 MBs, as well as other block types in the depth map coding, are provided in Table 4. For depth map coding, the block types other than skip and 16×16 MBs mostly occur in the object edge with discontinuous depth values, the MVs among neighboring blocks have relatively large differences. Thus, for these block types, the texture video MV would be more accurate than the median PMV and can be selected as the optimal PMV for depth maps. Thereby, the coding bits for MV difference in depth maps are saved. Moreover, from Table 4, it can be seen that the proposed method can give better coding results than the MV-competition method.

Table 3 Sequences parameters.

Sequences	Resolution	Left view	Central view	Right view
Breakdancers	1024×768	view 02	view 03	view 04
Ballet	1024×768	view 02	view 03	view 04
Bookarrival	1024×768	view 10	view 09	view 08

Table 4 Coding efficiency of different test sequences.

Sequences	H.264/AVC		MV-competition		Proposed		Percentage of Texture MV(%)		Compared with MV-competition		Compared with H.264/AVC	
	PSNR (dB)	Bit rate (kbit/s)	PSNR (dB)	Bit rate (kbit/s)	PSNR (dB)	Bit rate (kbit/s)	Skip & 16×16 MBs	Other block types	BD PSNR (dB)	BD bit rate (%)	BD PSNR (dB)	BD bit rate (%)
	Breakdancers view 02	47.74	707.32	47.72	705.24	47.74	702.22	0.84	27.34	0.05	-0.91	0.15
	43.99	372.83	43.95	365.7	43.96	364.63	1.85	29.88				
	40.64	189.9	40.63	184.52	40.63	181.02	2.57	32.38				
	37.23	93.68	37.29	90.84	37.26	90.44	2.81	33.86				
Breakdancers view 04	47.68	705.43	47.67	700.03	47.64	699.25	0.88	27.66	0.03	-0.74	0.13	-2.56
	43.91	364.97	43.89	360.93	43.92	361.02	1.75	30.18				
	40.57	180.89	40.61	177.07	40.60	175.20	1.91	32.49				
	37.1	86.45	37.1	83.43	37.15	82.60	2.06	33.95				
Ballet view 02	47.55	605.08	47.57	602.25	47.58	601.60	1.02	29.10	0.10	-1.50	0.24	-3.94
	43.98	349.71	43.95	343.69	43.89	341.66	1.95	34.52				
	39.94	184	39.97	179.68	39.95	172.90	2.06	40.43				
	35.73	80.62	35.81	77.35	35.72	74.92	2.85	44.19				
Ballet view 04	47.9	561.69	47.93	551.83	47.92	551.53	1.05	30.85	0.05	-1.09	0.20	-3.52
	44.23	337.75	44.24	330.55	44.22	328.68	1.44	36.70				
	40.16	178.2	40.17	173.85	40.15	172.37	1.75	42.03				
	36.17	78.53	36.23	77.64	36.17	71.56	2.06	44.04				
Bookarrival view 08	45.89	336.48	45.85	329.25	45.87	327.02	1.14	33.72	0.06	-1.34	0.24	-4.89
	42.68	177.13	42.7	170.2	42.69	169.27	1.75	37.60				
	39.77	95.05	39.76	91.62	39.72	88.36	1.87	39.91				
	36.69	52.49	36.74	50.75	36.65	49.67	2.74	41.39				
Bookarrival view 10	46.1	461.51	46.13	458.36	46.11	457.77	1.81	30.81	0.06	-1.30	0.21	-4.39
	42.87	238.03	42.93	237.07	42.89	230.65	2.01	36.33				
	39.85	122.18	39.84	114.83	39.81	113.38	2.95	39.44				
	36.6	62.18	36.6	61.54	36.58	59.28	3.05	41.48				
Average									0.06	-1.15	0.19	-3.68

To further verify the effectiveness of the proposed method, the qualities of the synthesized virtual views were also compared as shown in Table 5. As seen from Table 5, the average BD PSNR of synthesized virtual views using the proposed method was improved up to 0.13 dB compared with the median PMV utilized in H.264/AVC.

5 Conclusions

Based on the analyses of motion correlations between texture videos and depth maps, the texture video MV can be used as a candidate PMV for depth map coding in the proposed

method. Experimental results have verified that the proposed method can achieve higher coding efficiency than the median PMV utilized in H.264/AVC.

Acknowledgments

The work was supported by Xidian-ZTE Research Funds.

References

1. P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.* **17**(11), 1461–1473 (2007).
2. S. Grewatch and E. Müller, "Sharing of motion vectors in 3D video coding," in *Proc. ICIP*, Vol. 5, pp. 3271–3274, Singapore (2004).
3. H. Oh and Y. S. Ho, "H.264-based depth map sequence coding using motion information of corresponding texture video," in *Proceedings of PSIVT*, pp. 898–907, Hsinchu, Taiwan (2006).
4. H. Yuan, Y. Chang, Z. Lu, and Y. Ma, "Model based motion vector predictor for zoom motion," *IEEE Signal Process. Lett.* **17**(9), 787–790 (2010).
5. G. Laroche, J. Jung, and B. P. Popescu, "RD optimized coding for motion vector predictor selection," *IEEE Trans. Circuits Syst. Video Technol.* **18**(12), 1681–1691 (2008).
6. G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16 Q.6 Document, VCEG-M33, Austin, Texas (2001).

Table 5 PSNR comparison of the synthesized virtual view.

Sequences	BD PSNR (dB)	Average (dB)
Breakdancers virtual view 03	0.08	
Ballet virtual view 03	0.11	0.13
Bookarrival virtual view 09	0.20	