Rate-Distortion Optimization with Inter-View Refreshment for Stereoscopic Video Coding over Error-Prone Networks

Xinguang Xiang*[a], Debin Zhao[a], Qiang Wang[a], Siwei Ma[b], Wen Gao[b]

[a]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, P.R. China

[b]Institute of Digital Media, Peking University, Beijing, 100871, P.R. China

ABSTRACT

Stereoscopic video is a practical and important manner for 3-D video applications, and robust stereoscopic video transmission over error-prone networks has posed a technical challenge for stereoscopic video coding. In this paper, we present a rate-distortion optimization algorithm with inter-view refreshment for error-resilient stereoscopic video coding. First, inter-view refreshment is proposed to suppress error propagations besides intra refreshment. Then, we propose an end-to-end distortion model for stereoscopic video coding which concurrently considers network conditions, inter-view refreshment, and error concealment tools. Finally, based on the proposed end-to-end distortion model, a rate-distortion optimization algorithm is presented to adaptively select inter-view, inter and intra coding modes for error-resilient stereoscopic video coding. Simulation results show that the proposed scheme has a superior transmission efficiency improvement for stereoscopic video coding.

Keywords: Stereoscopic video coding, error resilience, inter-view refreshment, end-to-end distortion model, rate-distortion optimization

1. INTRODUCTION

Recent years have seen increased research efforts for stereoscopic video coding, because stereoscopic video, which can provide viewers the 3-D sensation, has been a practical and important approach for 3-D video applications. The basic prediction structure of stereoscopic video coding including disparity compensation prediction (DCP) [1] and motion compensation prediction (MCP) is shown in Fig. 1. In this paper, we put emphasis on the two-view based stereoscopic video, and the research can be easily extended from two views to cases of more than two views such as multi-view video.

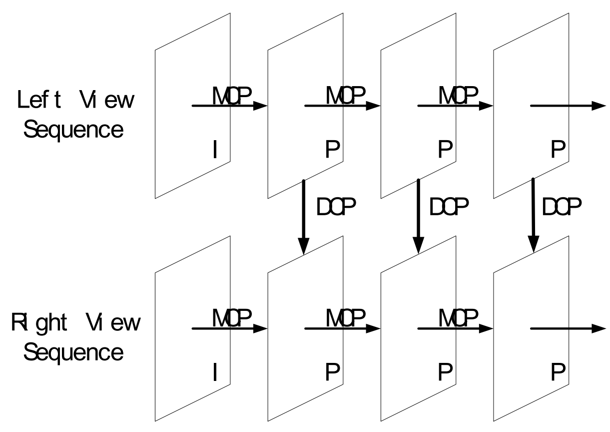

Fig. 1. Basic prediction structure of stereoscopic video coding

Robust video transmission over error-prone networks has posed a technical challenge for video coding. Error-resilient video coding which is executed at the encoder side and error concealment which is executed at the decoder side are two major categories of techniques to deal with the challenge. The main purpose of error-resilient video coding is to suppress

Visual Communications and Image Processing 2009, edited by Majid Rabbani, Robert L. Stevenson, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7257, 72570K · © 2009 SPIE-IS&T · CCC code: 0277-786X/08/$18 · doi: 10.1117/12.805643

error propagations caused by transmission errors. Inserting more intra-coded macroblocks is an effective technique to suppress error propagations but with the penalty on coding efficiency. Therefore, how to adaptively insert intra blocks to achieve better trade-off between coding efficiency and suppression of error propagations is an important issue. Adaptive intra/inter coding mode selection in the traditional single view coding has been widely exploited. The earliest algorithm which has been very popular and widely accepted is random intra update (RIU) [2]. Also, the content-adaptive coding mode selection method is proposed to insert intra blocks [3]. In [4], an end-to-end approach is proposed for error-resilient video coding. Then, several rate-distortion optimization algorithms [5][6] have been proposed for coding mode selection. These techniques are designed for single-view video coding, so they will be inefficient if directly applied in stereoscopic video coding, as the inter-view correlation is not considered at all.

To our knowledge, although a number of works [7][8] have been reported on error concealment of stereoscopic video coding, how to make adaptive coding mode selection for stereoscopic video coding over error-prone networks is still not fully researched. In this paper, we first propose that inter-view refreshment should be used for stereoscopic video coding as well as intra refreshment. Then, an end-to-end distortion model for stereoscopic video coding is proposed which concurrently considers network conditions, inter-view refreshment, and error concealment tools. Finally, we present a rate-distortion optimization algorithm which can adaptively select inter-view, inter and intra coding modes based on the proposed end-to-end distortion model for error-resilient stereoscopic video coding.

The rest of this paper is organized as follows. Section 2 describes the proposed algorithm in detail. Section 3 reports the simulation results. Section 4 concludes this paper.

2. THE PROPOSED RATE-DISTORTION OPTIMIZATION ALGORITHM FOR STEREOSCOPIC VIDEO CODING

2.1 Inter-view refreshment

Although intra coding can effectively suppress error propagations, it brings a big coding efficiency loss. We find that in stereoscopic video coding inter-view refreshment can provide a better trade-off between coding efficiency and suppression of error propagations. First, inter-view refreshment can provide similar error suppression as intra refreshment at some conditions. In stereoscopic video coding, DCP is widely used to exploit the inter-view correlation between the left and right views, so blocks in the right view may be predicted from the left view with disparity vectors. If the reference blocks in the left view are transmitted correctly, then the inter-view predicted blocks in the right view can be protected from error propagations of the right view. Furthermore, when the network condition of the left view is better than that of the right view or some unequal error protection techniques are utilized to make the packet loss rate of the left view lower than that of the right view, it can be expected that the left view will has lesser error propagations, where inter-view refreshment will suppress error propagations effectively. Even through the network condition of the left view is very bad, if the distortion distribution of the stereoscopic video can be estimated by an appropriate end-to-end distortion estimation algorithm, some contents of the right view video are also suitable for DCP to suppress error propagation. Therefore, improving the proportion of DCP also can make the right view robust to errors. Second, inter-view refreshment is usually more efficient than intra refreshment in terms of coding efficiency because of the inter-view correlation existing in stereoscopic video.

Based on the above analysis, we believe that inter-view refreshment is an effective error-resilient coding technique for stereoscopic video coding. Simulation results in section 3 will confirm this expectation.

2.2 The impact of error-prone network conditions on stereoscopic video coding

Stereoscopic video coding should adapt to error-prone network conditions to achieve better error resilience performance. In contrast to traditional single view error-resilient video coding, not only the network conditions (e.g., packet loss rate) of the current view but also that of the other view should be considered in error-resilient stereoscopic video coding for the current view. In stereoscopic video coding, DCP can also bring error propagations like MCP, while appropriate inter-view refreshment can still improve error resilience performance. When the network condition of the left view is better than that of the right view or some unequal error protection techniques are utilized to make the packet loss rate of the left view lower than that of the right view (especially when the packet loss rate of the left view is very low), more inter-view refreshment should be encouraged to protect the right view from error propagations. When the network condition of the left view is worse, more intra coding blocks are encouraged for the left view and the DCP proportion for the right view should be reduced. When the network conditions of the two views are almost the same, smart decisions should be taken to adaptively select inter-view, inter, and intra coding which can get a better trade-off between coding efficiency and suppression of error propagations. A rate-distortion optimization algorithm and the associated end-to-end distortion model are needed to realize the above coding strategy.

2.3 End-to-end distortion model for stereoscopic video coding

In our proposed end-to-end distortion model for stereoscopic video coding, the network conditions for stereoscopic video, inter-view refreshment, and error concealment tools for stereoscopic video coding are all considered.

For pixel *i* in frame *n* at viewpoint *view* which references pixel *j* in frame $ref(view,n)$, let $f^i_{view,n}$ denote the original value of the pixel *i*, and let $\hat{f}^i_{view,n}$ and $\tilde{f}^i_{view,n}$ denote the reconstructed values in the encoder and decoder respectively. $ref(view,n)$ may come from the left view or the right view decided by *view*, *n*, and coding mode. Let $\hat{r}^i_{view,n}$ be the reconstructed residue in the encoder, thus $\hat{f}^i_{view,n} = \hat{f}^j_{ref(view,n)} + \hat{r}^i_{view,n}$. When the current pixel is lost, it copies from pixel k in frame $ec\_ref(view,n)$, where $ec\_ref(view,n)$ also may come from the left view or the right view decided by *view*, *n*, and the used error concealment algorithm. Let p_{view} be the transmission error rate for viewpoint *view*. Then we can represent $\tilde{f}^i_{view,n}$ as

$$\tilde{f}^i_{view,n} = \begin{cases} \tilde{f}^j_{ref(view,n)} + \hat{r}^i_{view,n} & w.p. \quad 1-p_{view} \\ \tilde{f}^k_{ec\_ref(view,n)} & w.p. \quad p_{view} \end{cases} \tag{1}$$

Further, we can derive our end-to-end distortion $d(view,n,i)$ for stereoscopic video as

$$\begin{aligned} d(view,n,i) &= E\left\{\left(f^i_{view,n} - \tilde{f}^i_{view,n}\right)^2\right\} \\ &= (1-p_{view})E\left\{\left(f^i_{view,n} - \left(\tilde{f}^j_{ref(view,n)} + \hat{r}^i_{view,n}\right)\right)^2\right\} \\ &\quad + p_{view}E\left\{\left(f^i_{view,n} - \tilde{f}^k_{ec\_ref(view,n)}\right)^2\right\} \\ &\approx (1-p_{view})E\left\{\left(f^i_{view,n} - \hat{f}^i_{view,n}\right)^2\right\} + (1-p_{view})E\left\{\left(\hat{f}^j_{ref(view,n)} - \tilde{f}^j_{ref(view,n)}\right)^2\right\} \\ &\quad + p_{view}E\left\{\left(f^i_{view,n} - \tilde{f}^k_{ec\_ref(view,n)}\right)^2\right\} \\ &= (1-p_{view})d_s(view,n,i) + (1-p_{view})d_{ep}(ref(view,n),j) \\ &\quad + p_{view}d_{ec}\left(ec\_ref(view,n),i\right) \end{aligned} \tag{2}$$

where $d_s(view,n,i)$, $d_{ep}(ref(view,n),j)$, and $d_{ec}\left(ec\_ref(view,n),i\right)$ denote the source distortion, the error-propagation distortion from the reference frame, and the error concealment distortion, respectively. Obviously, the source distortion

$d_s(view,n,i)$ can be estimated at the encoder side, so we need to calculate the error-propagation distortion $d_{ep}(ref(view,n),j)$ and the error concealment distortion $d_{ec}\left(ec\_ref(view,n),i\right)$.

The error-propagation distortion from the reference frame $d_{ep}(ref(view,n),j)$ can be derived as

$$\begin{aligned}
d_{ep}(view,n,i) &= E\left\{\left(\hat{f}^i_{view,n}-\tilde{f}^i_{view,n}\right)^2\right\} \\
&= (1-p_{view})E\left\{\left(\hat{f}^i_{view,n}-\left(\tilde{f}^j_{ref(view,n)}+\hat{r}^i_{view,n}\right)\right)^2\right\}+p_{view}E\left\{\left(\hat{f}^i_{view,n}-\tilde{f}^k_{ec_ref(view,n)}\right)^2\right\} \\
&= (1-p_{view})E\left\{\left(\hat{f}^j_{ref(view,n)}-\tilde{f}^j_{ref(view,n)}\right)^2\right\} \\
&\quad +p_{view}E\left\{\left(\hat{f}^i_{view,n}-\hat{f}^k_{ec_ref(view,n)}+\hat{f}^k_{ec_ref(view,n)}-\tilde{f}^k_{ec_ref(view,n)}\right)^2\right\} \\
&\approx (1-p_{view})E\left\{\left(\hat{f}^j_{ref(view,n)}-\tilde{f}^j_{ref(view,n)}\right)^2\right\} \\
&\quad +p_{view}E\left\{\left(\hat{f}^i_{view,n}-\hat{f}^k_{ec_ref(view,n)}\right)^2\right\}+p_{view}E\left\{\left(\hat{f}^k_{ec_ref(view,n)}-\tilde{f}^k_{ec_ref(view,n)}\right)^2\right\} \\
&= (1-p_{view})d_{ep}(ref(view,n),j)+p_{view}d_{ec_r}\left(view,n,ec_ref(view,n),i\right) \\
&\quad +p_{view}d_{ep}(ec_ref(view,n),k)
\end{aligned} \tag{3}$$

where $d_{ec\_r}\left(view,n,ec\_ref(view,n),i\right)$ denotes the distortion between the reconstructed pixel values and the reconstructed error concealment pixel values in the encoder, which can be calculated at the encoder side. $d_{ep}(ref(view,n),j)$ can be derived from error-propagation distortions of its reference frame and its error-concealment reference frame. We utilize a block-level distortion map to store error-propagation distortions of frames in each view. Error-propagation distortion of the first frame in each view is set to be 0, because they are usually intra frames. Then the error-propagation distortions of the following frames can be calculated frame by frame by formula (3).

The error concealment distortion $d_{ec}\left(ec\_ref(view,n),i\right)$ can be derived as

$$\begin{aligned}
d_{ec}\left(view,n,i\right) &= E\left\{\left(f^i_{view,n}-\tilde{f}^k_{ec_ref(view,n)}\right)^2\right\} \\
&= E\left\{\left(f^i_{view,n}-\hat{f}^k_{ec_ref(view,n)}+\hat{f}^k_{ec_ref(view,n)}-\tilde{f}^k_{ec_ref(view,n)}\right)^2\right\} \\
&\approx E\left\{\left(f^i_{view,n}-\hat{f}^k_{ec_ref(view,n)}\right)^2\right\}+E\left\{\left(\hat{f}^k_{ec_ref(view,n)}-\tilde{f}^k_{ec_ref(view,n)}\right)^2\right\} \\
&= d_{ec_o}(view,n,ec_ref(view,n),i)+d_{ep}(ec_ref(view,n),k)
\end{aligned} \tag{4}$$

where $d_{ec\_o}(view, n, ec\_ref(view, n), i)$ denotes the distortion between the original pixel values and the reconstructed error concealment pixel values in the encoder, which also can be calculated at the encoder side. $d_{ep}(ec\_ref(view, n), k)$ can be calculated by formula (3). Thus the error concealment distortion $d_{ec}\left(ec\_ref(view, n), i\right)$ can be calculated at the encoder side.

Note that the source distortion, the error-propagation distortion from the reference frame, and the error concealment distortion all can be calculated at the encoder side. Thus the proposed end-to-end distortion for stereoscopic video can be utilized for coding mode selection.

In the proposed end-to-end distortion model for stereoscopic video coding, the error-propagation distortion caused by DCP is considered, thus inter-view refreshment can be used for error-resilient stereoscopic video coding. Also, the model considers network conditions of stereoscopic video, i.e., different packet-loss rates for two views. Furthermore, our proposed model can include error concealment algorithms for stereoscopic video coding in the distortion calculation.

2.4 Rate-distortion optimization for stereoscopic video coding

Based on the above end-to-end distortion model which concurrently considers network conditions, inter-view refreshment, and error concealment tools, we propose a rate-distortion optimization algorithm to adaptively select inter-view, inter and intra coding modes for stereoscopic video coding over error-phone networks.

2.4.1 Lagrange multiplier

The Lagrange multiplier indicates the relations between the distortion and the rate. In this subsection, we derive the new Lagrange multiplier for stereoscopic video coding over error-prone networks, considering both the packet loss distortion and the prediction structure of stereoscopic video coding.

Based on the derivation of Lagrange multiplier for single-view hybrid video coding [9], the source probability distribution can be approximated as uniform within each quantization interval, i.e., for each quantization interval Δ, the source distortion $D_s\left(\Delta\right)$ conforms to

$$D_s(\Delta) = \frac{\Delta^2}{12} , \tag{5}$$

and the rate $R\left(\Delta\right)$ conforms to

$$R\left(\Delta\right) = \frac{1}{\alpha}\log_2\left(\frac{\beta}{\Delta^2/12}\right) \tag{6}$$

where β denotes a constant depending on the variance of the source.

According to (2) and (5), we can obtain

$$D\left(\Delta\right) = \left(1 - p_{view}\right)\frac{\Delta^2}{12} + \left(1 - p_{view}\right)D_{ep} + p_{view}D_{ec} \tag{7}$$

where the error-propagation distortion D_{ep} and the error concealment distortion D_{ec} are both independent of the quantization interval Δ of current frame.

Combining (6) and (7), we can derive the new Lagrange multiplier

$$\lambda_{view} = -\frac{dD(R)}{dR} = -\frac{dD}{d\Delta}\frac{d\Delta}{dR} = (1-p_{view})\frac{\alpha \ln 2}{12}\Delta^2 = (1-p_{view})\lambda_0 \tag{8}$$

for stereoscopic video coding, where p_{view} denotes the packet loss rate and λ_0 is the Lagrange multiplier used in H.264 or H.263. Notice that each view has its own λ_{view} decided by its packet loss rate p_{view}.

2.4.2 Mode selection for error-resilient stereoscopic video coding

The proposed rate-distortion optimization algorithm utilizes the above end-to-end distortion model and the Lagrange multiplier to select the optimal coding mode for error-resilient stereoscopic video coding. According to (2) and (8), supposing m_J denotes motion vectors for the macroblock m, the coding mode $o^*(view, n, m)$ can be decided by

$$\begin{aligned} o^*(view,n,m) &= \underset{o\in O}{\arg\min}\Big((1-p_{view})\Big(D_s(view,n,m,o)+D_{ep}\big(ref(view,n),m_J\big)\Big) \\ &\quad + p_{view}D_{ec}\big(er\_ref(view,n),m\big)+\lambda_{view}R\Big) \\ &= \underset{o\in O}{\arg\min}\Big(\Big(D_s(view,n,m,o)+D_{ep}\big(ref(view,n),m_J\big)\Big)+\lambda_0 R\Big) \end{aligned} \tag{9}$$

where O denotes the set of all candidate coding modes including intra, inter and inter-view coding.

The proposed rate-distortion optimization algorithm considers not only the source distortion, but also the error-propagation distortion. Thus the proposed algorithm can adaptively select inter-view, inter and intra coding modes for error-resilient stereoscopic video coding.

3. SIMULATION RESULTS

The H.264 reference software JM 10.0 is used to simulate a stereoscopic video coding system like Fig. 1. In the system, we utilize the error concealment algorithm for stereoscopic video coding proposed in our former work [8], because the inter-view correlation in stereoscopic videos is effectively exploited in [8]. This system serves as anchor, while the random intra update (RIU) algorithm and our proposed algorithm (PR) are simulated on the system. The sequences Race1 and Ballroom are encoded using IPPP GOP structure for 225 frames and the packet loss rates (PLR) at 5%, 10% and 20% are tested.

First, in order to evaluate the performance of the proposed algorithm with inter-view refreshment on the right view, which is predicted by both DCP and MCP, we assume that the left view sequence is error free and the right view sequence is transmitted with packet loss. The PSNR values of the right view at the decoder side under the given conditions are shown in Table 1. The proposed algorithm has a 0.87~2.68dB transmission performance improvement than the anchor and a 0.63~1.5dB improvement than RIU. Table 2 shows the average bit-rate savings of proposed algorithm compared to RIU with PLR=20%. Compared to RIU, the proposed algorithm can have a 17.22%~20.82% average bit-rate savings. The results indicate that the proposed algorithm with inter-view refreshment is very efficient for stereoscopic video coding.

In Table 3, we can see the decoded PSNR values of the stereoscopic pairs of sequence when both the left and right view bitstreams are transmitted with packet loss. Both the case that two views have the same PLR and the case that two views have different PLRs are tested. The proposed algorithm has a 1.36~4.05dB transmission performance improvement than the anchor and a 0.73~2.39dB improvement than RIU. The results show that the proposed rate-distortion optimization algorithm with inter-view refreshment provides superior transmission efficiency for error-resilient stereoscopic video coding.

Table 1 Average PSNR (dB) comparison for the right view sequence supposing the left view is error free

Video sequence	Scheme	Packet loss rates		
		5%	10%	20%
Race1	Anchor	34.67	32.58	30.76
	RIU	35.69	33.85	31.94
	PR	36.55	35.16	33.44
Ballroom	Anchor	34.11	32.33	29.95
	RIU	34.35	32.76	30.93
	PR	34.98	33.95	31.80

Table 2 Average bit-rate savings of the proposed algorithm compared to RIU

Video sequence	Bit-rate savings
Race1	17.22%
Ballroom	20.82%

Table 3 Average PSNR (dB) comparison for stereoscopic video at different packet loss rates

Video sequence	Scheme	Packet loss rates								
		Left view: 5%	5%	5%	10%	10%	10%	20%	20%	20%
		Right view: 5%	10%	20%	5%	10%	20%	5%	10%	20%
Race1	Anchor	33.51	32.60	31.76	31.58	30.93	30.19	29.83	29.29	28.76
	RIU	33.76	32.97	32.60	32.23	31.69	31.43	31.27	30.75	30.61
	PR	34.87	34.23	33.48	34.01	33.34	32.53	33.14	32.38	31.34
Ballroom	Anchor	32.46	31.77	30.86	30.69	30.20	29.43	28.44	28.14	27.60
	RIU	32.92	32.37	31.73	31.66	31.30	30.74	30.10	29.90	29.58
	PR	34.72	33.89	32.85	33.83	32.94	31.86	32.49	31.59	30.41

4. CONCLUSIONS

Inter-view correlation in stereoscopic video is useful for error-resilient stereoscopic video coding. In this paper, a rate-distortion optimization algorithm with inter-view refreshment is proposed for error-resilient stereoscopic video coding. First, we propose that inter-view refreshment is an effective technique to suppress error propagations besides intra refreshment. Further, an end-to-end distortion model for stereoscopic video coding is proposed, taking into account network conditions, inter-view refreshment and error concealment tools for stereoscopic video. Finally, based on the proposed end-to-end distortion model, we present a rate-distortion optimization algorithm which can adaptively select inter-view, inter and intra coding modes for error-resilient stereoscopic video coding. Simulation results show that the proposed scheme provides superior transmission efficiency for stereoscopic video coding. The scheme can be easily extended from two views to cases of more than two views such as multi-view video coding.

5. ACKNOWLEDGMENT

This work was supported in part by National Science Foundation (60672088 and 60736043) and Major State Basic Research Development Program of China (973 Program, 2009CB320905).

REFERENCES

[1] M. E. Lukacs, "Predictive coding of multi-viewpoint image sets," Proc. IEEE ICASSP 86, pp. 521-524, 1986.

[2] G. Cote and F. Kossentini, "Optimal intra coding of blocks for robust video communication over the internet," Signal Process.: Image Commun., vol.15, pp. 25-34, Sept. 1999.

[3] P. Haskell and D. Messerschmitt, "Resynchronization of motion-compensated video affected by ATM cell loss," Proc. IEEE ICASSP 92, vol.3, pp. 545-548 (1992).

[4] D. Wu, Y. T. Hou, B. Li, W. Zhu, Y. Q. Zhang, and H. J. Chao, "An end-to-end approach for optimal mode selection in Internet video communication: theory and application," IEEE J. Select. Areas Commun., vol.18, no.6, pp. 977-995, Jun. 2000.

[5] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packer loss resilience," IEEE J. Select. Areas Commun., vol.18, no.6, pp. 966-976, Jun. 2000.

[6] T. Stockhammer, D. Kontopodis, and T. Wiegand, "Rate-distortion optimization for JVT/H.26L coding in packet loss environment," Proc. Packet Video Workshop, Pittsburgh, PA, Apr. 2002.

[7] S. Knorr, C. Clemens, M. Kunter and T. Sikora, "Robust concealment for erroneous block bursts in stereoscopic images", Proc. of 3DPTV2004, pp.820-827, Sept. 2004.

[8] X. Xiang, D. Zhao, Q. Wang, X. Ji, and W. Gao, "A novel error concealment method for stereoscopic video coding," Proc. IEEE ICIP2007, pp.101-104, Sep. 2007.

[9] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint source-channel rate-distortion optimization for H.264 video coding over error-prone networks," IEEE Transactions on Multimedia, vol.9, no.3, pp. 445-454, Mar. 2007.