ESTIMATION OF END-TO-END DISTORTION OF VIRTUAL VIEW FOR ERROR-RESILIENT DEPTH MAP CODING

Min Gao, Xiaopeng Fan, Debin Zhao, Feng Jiang, Wen Gao

Dept. of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
{mgao, fxp, dbzhao and fjiang }@hit.edu.cn, wgao@jdl.ac.cn

ABSTRACT

In this paper, we propose to estimate the end-to-end distortion of virtual view for error-resilient depth map coding by analyzing the view-rendering process. The end-to-end distortion is estimated according to the expectation of depth distortion. The expectation of depth distortion is estimated based on the error-propagated distortion from the reference frame, which is calculated recursively for each pixel by considering the source characteristics, channel condition and error concealment method. The error-propagated distortion in terms of each frame is updated after the frame is encoded, which is used for the subsequent frames. Experimental results demonstrate that the proposed algorithm achieves substantial and consistent gains in comparison with the random intra updating method and the conventional rate-distortion method for error-resilient depth map coding.

Index Terms— Depth map coding, error-resilient coding, end-to-end distortion, virtual view

1. INTRODUCTION

Video representations based on view synthesis using depth map, such as multi-view plus depth (MVD), have been recently proposed. In these video representations, depth map is gray-scale video, which contains the depth information of the corresponding pixel in color video. Usually, the size of raw depth map is typically one third of raw color video. To reduce the data amount of depth map, it is commonly compressed by the conventional video coding standards, such as H.264/AVC or MVC.

When the compressed depth map is transmitted through error-prone networks, robustness to packet loss is a crucial requirement, because the prediction loop in the conventional video coding standards can propagate errors and cause substantial degradation of depth map quality.

To deal with this error propagated problem, insertion of intra coded blocks is a widely used technique. On one hand, intra coding will suppress error propagation. However, on the other hand, it will reduce the coding efficiency. Therefore, how to insert intra coded blocks in terms of rate and distortion is necessary.

There have been some intra refreshment algorithms for error resilient coding of color video. An algorithm to randomly insert the intra coded block was proposed in [1]. To further improve the performance, several rate distortion optimized methods have been proposed, in which the end-to-end distortion of color video was estimated. In [2], a recursive optimal per-pixel estimate (ROPE) algorithm has been proposed to estimate the end-to-end distortion at pixel level by tracking the first and second moments of the reconstructed pixel value. An error-resilient rate distortion model adopted in H.264/AVC test model has been proposed in [3], in which the expected end-to-end distortion is estimated in a manner of independently simulating K copies of channel behaviors at the encoder. The end-to-end distortion model in [4] takes the overall distortion as the sum of several separate distortion items.

The above methods can enhance the error robustness of color video to packet loss. However, they are not suitable for depth map, because depth map is not displayed to users and it is only a supplement data to synthesize the virtual view, which is displayed to users. To improve coding efficiency of source coding for depth map in error free environment, some distortion metrics have been proposed. Kim *et al* in [5] [6] analyzed the impact of depth distortion for virtual view quality and proposed a new distortion metric. To improve the accuracy of the distortion model, the disparity rounding problem was considered in [7]. An alternative model has been proposed in [8] by mimicking the view rendering process. However, these distortion models are not suitable for error-resilient depth map coding since the channel condition and error concealment method are not considered.

In this paper, we propose to estimate the end-to-end distortion of virtual view for error-resilient depth map coding by analyzing the view-rendering process. The end-to-end distortion is estimated according to the expectation of depth distortion. The expectation of depth distortion is estimated based on the error-propagated distortion from the reference frame, which is calculated recursively for each pixel by considering the source characteristics, channel condition and error concealment method.

The rest of the paper is organized as follows. Section 2 analyzes the virtual view distortion. Section 3 describes the proposed algorithm. In Section 4, experimental results are presented. Finally, Section 5 concludes the paper.

2. VIRTUAL VIEW DISTORTION

In this section, we analyze the virtual view distortion function in [8] to show the impact of depth error to virtual view quality by mimicking the warping process.

If there are a bunch of color and depth pixels, $C_1, \ldots, C_k$ and $D_1, \ldots, D_k$, then each color pixel C_k will be warped to $C_{k'}$ by the collocated depth pixel D_k. After warping, the interpolated pixel at integer grid will be displayed if the warped position k' is not integer. For simplification, it is assumed that only the depth of the current pixel D_k becomes noisy to $\widehat{D}_k$ and the linear interpolation scheme is used to interpolate the integer pixel. Hence, the noise-free warping curve A(x) is distorted to warping curve B(x) by the distortion of the current depth pixel D_k, as shown in Fig.1. Here, it is worthwhile to note that depth error will only change the warping position along *x*-direction and the color pixel value will only affect the intensity of the warped pixel along *y*-direction.

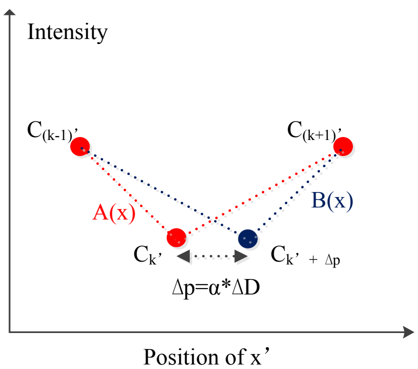


Fig.1. Illustration of virtual view distortion caused by depth distortion

Therefore, the distortion of the virtual view caused by the distortion of the current depth pixel will be the area between two curves A(x) and B(x). It can be approximated by the sum of area of two triangles, which is.

$$D(k) \cong 0.5 \cdot b \cdot (h_1 + h_2)$$
$$= 0.5 \cdot \Delta p \cdot [|C_k - C_{k-1}| + |C_k - C_{k+1}|] \quad (1)$$

where $\Delta p = \alpha \cdot \Delta D$, $\Delta D = |D_k - \widehat{D}_k|$ and α is proportional coefficient determined by the following equation.

$$\alpha = \frac{f \cdot L}{255} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right) \quad (2)$$

where f is focal length, L is baseline between current view and virtual view, and Z_{near} and Z_{far} are the values of the nearest and farthest depth of the scene, respectively.

3. THE PROPOSED ESTIMATION OF END-TO-END DISTORTION OF VIRTUAL VIEW

The proposed algorithm estimates the end-to-end distortion of virtual view according to the expectation of depth distortion according to the description in Section 2. In the following, we firstly analyze the expectation of depth distortion by considering two cases depending on whether the pixel belongs to intra coded block or inter coded block, and then introduce a method to estimate it, which is similar to our previous work in [4] used for color video.

3.1 Pixel in intra-coded block

Let D_n^i be the original depth of pixel *i* in frame *n*, and $\widehat{D}_n^i$ denote its reconstruction at encoder. The reconstructed depth at decoder is denoted as $\widetilde{D}_n^i$, which can be treated as a random variable at encoder due to the random channel behavior. *p* is the packet loss rate. If the packet containing intra coded block to which pixel *i* belongs is correctly received, then $\widetilde{D}_n^i = \widehat{D}_n^i$ and the probability of this event is 1- *p*. If the packet is lost, the depth of pixel *k* in the previous frame *n*-1 is used to estimate the pixel *i* in the current frame, then $\widetilde{D}_n^i = \widetilde{D}_{n-1}^k$ and the probability of this event is *p*. Therefore, the expectation of depth distortion can be derived as follows:

$$E\{|D_n^i - \widetilde{D}_n^i|\} = (1-p) \cdot |D_n^i - \widehat{D}_n^i| + p \cdot E\{|D_n^i - \widetilde{D}_{n-1}^k|\}$$
$$= (1-p) \cdot d_s(n,i) + p \cdot d_{ec}(n,i) \quad (3)$$

where $d_s(n,i)$ is the distortion caused by quantization and $d_{ec}(n,i)$ is the distortion caused by error-concealment.

3.2 Pixel in inter-coded block

Let us assume that pixel *i* in the current frame *n* is predicted from pixel *j* in the reference frame *ref* and the encoder prediction is $\widehat{D}_{ref}^j$. We denote the encoder quantization residue as $\hat{e}_n^i$, which means that the reconstruction depth of this pixel at encoder, $\widehat{D}_n^i$, is equal to $\widehat{D}_{ref}^j + \hat{e}_n^i$.

If the packet consisting of quantization residue and motion vector is correctly received, the decoder reconstruction of depth of pixel *i* in frame *n* is given by $\widetilde{D}_n^i = \hat{e}_n^i + \widetilde{D}_{ref}^j$, which is potentially different from the encoder reconstruction $\widehat{D}_n^i$ due to the difference between $\widetilde{D}_{ref}^j$ and $\widehat{D}_{ref}^j$. The probability of this event is 1-*p*. If the packet is lost, the reconstruction depth of pixel *k* in the previous frame *n-1* at decoder is used to estimate the reconstruction of pixel *i* in the current frame, that is $\widetilde{D}_n^i = \widetilde{D}_{n-1}^k$ and the probability of this event is *p*. Therefore the expectation of depth distortion can be computed by (4) based on the assumption that the effects of quantization and error-propagation are additive. In (4), $d_s(n,i)$ is the distortion caused by quantization, $d_{ep}(ref,j)$ is the distortion caused by error-propagation and $d_{ec}(n,i)$ is the distortion caused by error concealment.

$$
\begin{aligned}
E\{|D_n^i - \widetilde{D}_n^i|\} &= (1-p)\cdot E\{|D_n^i - (\hat{e}_n^i + \widetilde{D}_{ref}^j)|\} \\
&\quad + p\cdot E\{|D_n^i - \widetilde{D}_{n-1}^k|\} \\
&= (1-p)\cdot \mathrm{E}\{|D_n^i - \widehat{D}_n^i + \widehat{D}_{ref}^j - \widetilde{D}_{ref}^j|\} \\
&\quad + p\cdot \mathrm{E}\{|D_n^i - \widetilde{D}_{n-1}^k|\} \\
&= (1-p)\cdot \mathrm{E}\{|D_n^i - \widehat{D}_n^i|\} \\
&\quad + (1-p)\cdot \mathrm{E}\{|\widehat{D}_{ref}^j - \widetilde{D}_{ref}^j|\} \\
&\quad + p\cdot \mathrm{E}\{|D_n^i - \widetilde{D}_{n-1}^k|\} \\
&= (1-p)\cdot d_s(n,i) + (1-p)\cdot d_{ep}(ref,j) \\
&\quad + p\cdot d_{ec}(n,i) \qquad (4)
\end{aligned}
$$

3.3 Estimation of expectation of depth distortion

According to the above analysis, it can be seen that the expectation of depth distortion is determined by quantization distortion d_s, error-propagation distortion d_{ep} and error concealment distortion d_{ec}. Since d_s can be obtained at encoder, the expectation of depth distortion mainly relies on d_{ep} and d_{ec}.

Firstly, we derive $d_{ec}(n,i)$ as follows based on the assumption that the effects of original error concealment at encoder and error-propagation at decoder are additive.

$$
\begin{aligned}
d_{ec}(n,i) &= E\{|D_n^i - \widetilde{D}_{n-1}^k|\} \\
&= E\{|D_n^i - \widehat{D}_{n-1}^k + \widehat{D}_{n-1}^k - \widetilde{D}_{n-1}^k|\} \\
&= \mathrm{E}\{|D_n^i - \widehat{D}_{n-1}^k|\} + \mathrm{E}\{|\widehat{D}_{n-1}^k - \widetilde{D}_{n-1}^k|\} \\
&= d_{ec\_o}(n,i) + d_{ep}(n-1,k) \qquad (5)
\end{aligned}
$$

where $d_{ec\_o}(n,i)$ is the difference between the original depth and the error concealment one at encoder, and $d_{ep}(n-1,k)$ is the distortion caused by error-propagation.

Now the problem is how to calculate $d_{ep}(n-1,k)$ since d_{ec_o} is also obtained at encoder.

Since $d_{ep}(n-1,k)$ in (5) is similar to $d_{ep}(ref,j)$ in (4), without losing generality, we derive $d_{ep}(n,i)$ as follows:

$$
\begin{aligned}
d_{ep}(n,i) &= E\{|\widehat{D}_n^i - \widetilde{D}_n^i|\} \\
&= (1-p)\cdot E\{|\widehat{D}_n^i - (\widetilde{D}_{ref}^j + \hat{e}_n^i)|\} \\
&\quad + p\cdot E\{|\widehat{D}_n^i - \widetilde{D}_{n-1}^k|\} \\
&= (1-p)\cdot E\{|\widehat{D}_{ref}^j - \widetilde{D}_{ref}^j|\} \\
&\quad + p\cdot E\{|\widehat{D}_n^i - \widehat{D}_{n-1}^k|\} + p\cdot E\{|\widehat{D}_{n-1}^k - \widetilde{D}_{n-1}^k|\} \\
&= (1-\mathrm{p})\cdot d_{ep}(ref,j) + p\cdot d_{ec\_r}(n,i) \\
&\quad + p\cdot d_{ep}(n-1,k) \qquad (6)
\end{aligned}
$$

where $d_{ec\_r}(n,i)$ is the difference between the reconstructed depth of the current frame *n* and the error concealment depth value at encoder. Since $d_{ec\_r}(n,i)$ is available at encoder, $d_{ep}(n,i)$ can be calculated recursively if that d_{ep} of the first frame can be directly set zero since it is typically encoded as an intra-frame.

According to (3), (4) and (5), the expectation of depth distortion of the current frame can be calculated by referencing the error-propagation distortion of the previous frame. When the current frame is encoded, its error-propagation distortion should be updated according to (6).

3.4 Estimation of end-to-end distortion of virtual view

If the expectation of depth distortion of pixel *i* in frame *n*, denoted as $d = E\{|D_n^i - \widetilde{D}_n^i|\}$, is obtained according to (3) and (4). The end-to-end distortion of virtual view caused by this pixel can be estimated by substituting ΔD in equation (1) for *d*, which is written as follows.

$$D(i)_{end\_to\_end} = 0.5\cdot\alpha\cdot d\cdot(|C_i - C_{i-1}| + |C_i - C_{i+1}|) \qquad (7)$$

where α and C_i are the same as that in Section 2. The distortion obtained by (7) can be used in RDO based mode selection process using the Lagrangian optimization, with Lagrangian cost *J* written as:

$$\mathrm{J} = \sum_{i=0}^{N-1} D(i)_{end\_to\_end}^2 + \lambda\cdot R_{depth} \qquad (8)$$

where *N* is the block size of the current depth block and R_{depth} is the bits consumed for coding the current depth block.

4. EXPERIMENTAL RESULTS

The proposed algorithm is incorporated within the rate-distortion framework in order to optimally choose the number and position of intra coded block. We adopted the random intra-updating algorithm (“RU”) and the conventional rate-distortion algorithm (“Con-RD”) as the competing methods. In random intra-updating algorithm, given the packet loss rate *p*, a fraction of *p* macro-blocks in each frame are coded as intra. In RU and Con-RD algorithms, the virtual view distortion model in Section 2 is adopted in rate-distortion framework, since the proposed algorithm is derived based on the distortion model in Section 2.

The JM 15.1 H.264/AVC codec was employed. In the experiment, we employed CABAC for entropy coding; a single reference frame is used in motion estimation and sub-pixel motion estimation is disabled. The de-blocking filter is enabled and inter pixels are not used for intra macro-block prediction. Each row of macro-blocks in each frame composes a slice and is packed in a separate packet. The original rate control algorithm from JM codec is used and the temporal-replacement is employed for error concealment.

In the experiment, the depth maps from different views are compressed independently under the same bit rate. The first frame in depth map sequence is encoded as I-frame, and all remaining frames are encoded as P-frame. The reconstructed depth map and the original texture video are used to generate the virtual view by view synthesis reference software (VSRS) version 3.5. The virtual view synthesized by uncompressed depth map and original texture video is used as reference to compute PSNR. A set of 200 randomly generated packet loss patterns were applied at each packet loss rate and the average luminance PSNR of the virtual view is computed to evaluate the performance.

As a test set, we selected five multi-view sequences *Kendo*, *Balloons*, *Newspaper*, *Lovebird1* and *BookArrival*.

The resolution of these sequences is 1024x768. The detailed test setting is shown in Table 1.

Table 1. Test setting for each sequence

	Kendo	Balloons	Newspaper	Lovebird1	BookArrival
Bit rate(*kbps*)	746.40	688.57	788.96	528.49	518.57
Frames	300	150	150	200	100
View-number	1-3	1-3	4-6	4-6	8-10
Frame rate(*fps*)	30	30	30	30	16.67

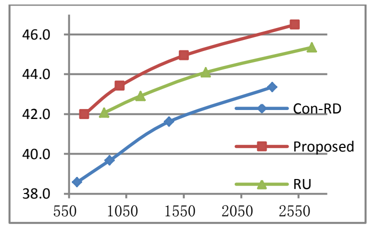
At the encoder, five bit streams are generated in terms of each algorithm; at the decoder, these bit streams are decoded after simulating packet loss rate 5%, 10%, 15% and 20%, respectively. It is assumed that the packet containing the parameter set and packets containing the first frame are correctly received at the decoder. The performance comparison for different sequences is presented in Table 2.

Table 2. Performance comparisons of average PSNR [dB] for virtual view at different packet loss rate

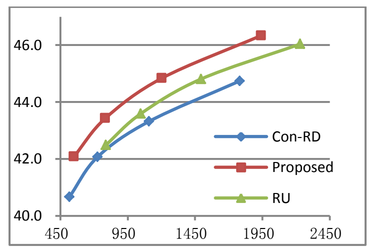
sequence	scheme	PSNR of different packet loss rate			
		5%	10%	15%	20%
Kendo	proposed	44.83	43.74	42.81	42.01
	RU	43.61	42.24	41.23	40.51
	Con-RD	41.47	38.75	36.97	35.67
Balloons	Proposed	45.13	44.58	44.05	43.67
	RU	44.48	43.49	42.74	42.14
	Con-RD	43.85	42.05	40.64	39.51
Newspaper	Proposed	41.86	41.17	40.61	40.02
	RU	41.16	39.96	39.12	38.41
	Con-RD	40.17	38.32	37.02	36.05
Lovebird1	Proposed	44.55	43.37	42.50	41.80
	RU	43.10	41.55	40.13	38.70
	Con-RD	43.71	42.16	40.99	40.05
BookArrival	Proposed	42.33	41.59	41.06	40.54
	RU	41.70	40.80	39.94	39.38
	Con-RD	41.16	39.58	38.34	37.31

From Table 2, it can be seen that the proposed algorithm outperforms the RU algorithm for all sequences. The RU algorithm outperforms the Con-RD algorithm except *Lovebird1*. The depth of the foreground and background are nearly not changed in sequence *Lovebird1*; based on the assumption that the first frame is correctly received, for some regions, it is not necessary to code them as intra, since the depth after error concealment approximates to the decoded one. Therefore, compared with the RU algorithm, the Con-RD algorithm can allocate more bits to the regions which have more influence on the quality of virtual view. Therefore, the Con-RD algorithm achieves higher quality than the RU algorithm for *Lovebird1*.

The rate-distortion curves for *Kendo* and *Balloons* at packet loss rate 5% are also shown in Fig. 2, in which *x*-direction is bitrate (Kbps) and *y*-direction is average Y-PSNR (dB). The QP values include 24, 28, 32 and 36. From Fig. 2, it can be seen that the proposed algorithm achieves better performance than RU and Con-RD methods.



(a) Kendo



(b) Balloons

Fig.2. Rate-distortion curves of the proposed method.

5. CONCLUSION

In this paper, an algorithm to estimate the end-to-end distortion of virtual view is proposed for error-resilient depth map coding. Different from the existing virtual view distortion functions used in the source coding of depth map in error-free environment, we take into account the channel condition and the error concealment algorithm used in the decoder. We incorporate the proposed algorithm into the rate-distortion framework to optimally choose the number and position of the intra coded blocks. Experimental results show that the proposed algorithm can achieve substantial and consistent gains compared with the random intra-updating algorithm.

ACKNOWLEDGEMENT

This work was supported in part by the Major State Basic Research Development Program of China's 973 Program under Grant 2009CB320905, the National Science Foundation of China under Grant 61100095 and 61100096, and the Program for New Century Excellent Talents in University (NCET) of China (NCET-11-0797).

REFERENCES

[1] G. Cote and F. Kossentini, “Optimal intra coding of blocks for robust video communication over the internet,” Signal Processing: Image Communication., vol. 15, pp. 25–34, Sep. 1999.

[2] R. Zhang, S. L. Regunathan, and K. Rose, “Video coding with optimal inter/intra-mode switching for packet loss resilience,” IEEE Journal on Selected Areas in Communications., vol. 18, pp. 966-976, Jun. 2000.

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

[4] 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 Trans. Multimedia, vol. 9, no. 3, pp. 445–454, Apr.2007.

[5] W. S. Kim, A. Ortega, P. L. Lai, D. Tian, and C. Gomila, “Depth map coding with distortion estimation of rendered view,” in Proc. SPIE Vis. Inf. Process. Commun., 2010.

[6] W. S. Kim, A. Ortega, P. L. Lai, D. Tian, and C. Gomila, “Depth map distortion analysis for view rendering and depth coding,” in Proc. IEEE Int. Conf. Image Process., 2009, pp. 721–724.

[7] T. Zhang, Xiaopeng Fan, Debin Zhao and Wen Gao, “New Distortion Model for Depth Coding in 3DVC,” in Proc. IEEE Vis. Inf. Process. Commun., 2012.

[8] B. T. Oh, J. Lee and D. Park, "Depth Map Coding Based on Synthesized View Distortion Function," IEEE Journal of Selected Topics in Signal Processing., vol. 5, pp. 1344-1352, Nov. 2011.