REDUCED RESOLUTION UPDATE FOR DEPTH MAP COMPRESSION ON THE BASIS OF TEXTURE AND PREDICTED IMAGE

Shiori Sugimoto,Shinya Shimizu,Hideaki Kimata, Akira Kojima

NTT Media Inteligence Laboratories, NTT Corporation

ABSTRACT

In 3D video compression, depth map is normally compressed by resolution reduction. To avoid artifacts in views synthesized with the reduced resolution depth map, up-sampling approaches based on special properties of depth map have been proposed. Most previous studies do not apply rate-distortion control and the quality of depth map and synthesized views is not always guaranteed because their re-sampling processes are out of the coding loop. This paper proposes a novel in-loop reduced resolution depth coding based on the scheme of a reduced resolution update mode. This is achieved by encoding prediction residuals at reduced resolution, which was obtained by performing prediction using full resolution references. We also propose an effective method for reduction of resolution for residuals and reconstruction of depth map based on Texture and predicted depth map. Experiments showed that it is possible to achieve bitrate reductions of up to 2.09% and about 1.10% on average for seven sequences.

Index Terms— 3D video coding, Depth map coding, Resolution reduction, Reduced resolution update, Depth reconstruction filter

1. INTRODUCTION

In recent years, 3D video has gained considerable attention in next generation media applications. With most past 3D display systems based on stereoscopic video data, representation of videos and their data volume is particularly unproblematic because stereoscopic video data is composed of only 2-view 2D video sequences at most. However, many organizations have developed auto-stereoscopic 3D displays to commercial product in the last few years. These need a multitude of videos, and their data volume increases in proportion to the number of views. To make matters worse, the required numbers of views and disparity values between views are absolutely display-dependent. Therefore, it may be necessary to synthesize desired video sequences at any viewpoint from fixed 3D video data [4].

One of the data formats developed in response to this requirement is known as view plus depth. In its cardinal form, 3D video is composed of multiple video sequences and their per-pixel depth information is represented as gray-scale video sequences. Given the depth information in an explicit form, desired views can be synthesized by a technique referred to as depth Image Based Rendering [3]. Therefore, the data of View plus depth is quite well represented and compressed as multiview video data. This gray-scale video sequence is known as depth map and regular color video sequence is called Texture. MPEG and VCEG have established a Joint Collaborative Team on 3D video coding to handle this format [5].

In compression of such 3D video data, depth map is often compressed with resolution reduction. This is caused by some special properties of depth map that are quite different from those of common 2D video data. One typical property is that the quality of depth map is often measured by the quality of synthesized views rendered on the basis of depth information instead of the depth map image quality itself. Another is that the depth information is very sensitive to the quantization noise [2]. Therefore, the bitrate of 3D video data was effectively decreased by reducing the resolution of depth map lower than the resolution of Texture. However, these reduction and reconstruction processes cannot apply rate-distortion control because the related resolution reduction approach for depth map compression consists of down-sampling as a pre-process of encoding and up-sampling as a post-process of decoding.

In this paper, we propose a novel approach of in-loop reduced resolution depth coding. The basic framework is based on reduced resolution update mode (RRU), which already exists within H.263[8]. It is a mechanism that can increase coding picture rate while maintaining sufficient subjective quality. This is achieved by encoding prediction residuals at reduced resolution, which was achieved by performing prediction using full resolution references. Our advanced scheme has a sub-sampling process for residual values in the same way that RRU has an encoding process, but it also has an up-sampling process for depth map values reconstructed from sub-sampled residual values and corresponding predictive values instead of up-sampling for residual values and a normal reconstruction process. Our scheme prevents noisy errors of reconstructed depth map caused by interpolation of residuals. We also propose a non-linear sub-sampling for residuals and corresponding up-sampling for reconstructed depth map. Experimental results obtained in software simulation show the effectiveness of the proposed approach.

2. RELATED STUDIES

This section introduces related approaches of reduced resolution depth coding and RRU.

2.1. Reduced Resolution Depth Coding

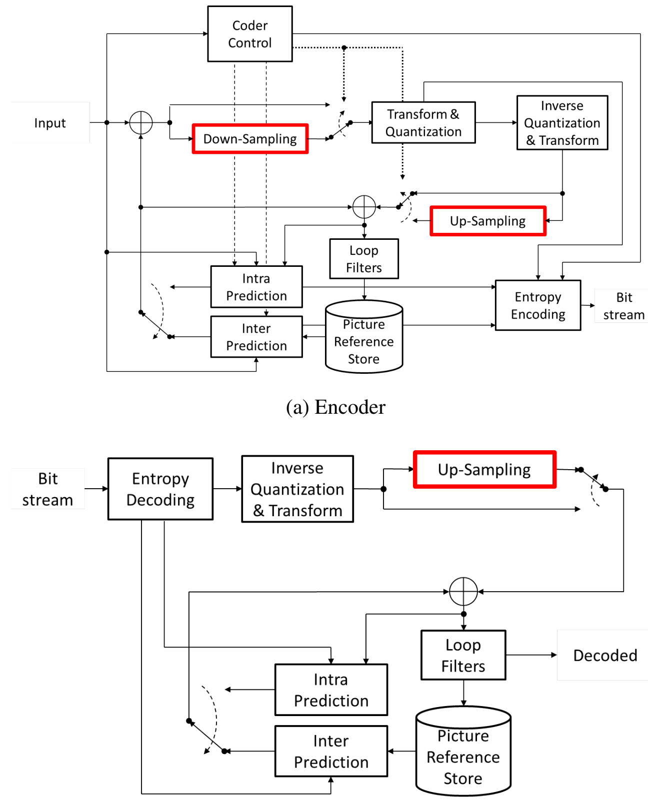
The effectiveness of reduced resolution depth coding is caused by special properties of depth map on the assumption that the quality of depth map is measured by the quality of rendered synthesized views based on depth information. Depth map contained in 3D video data is basically compressed as regular 2D video streams. An inevitable result is that the depth information obtained from coded depth map is affected by frequency based compression such as DCT. Coding noises such as mosquito noise cause strongly marked synthesis errors and it is difficult to restore such distorted depth information. On the other hand, because depth map has the property of piecewise smoothness, its spatial resolution can be reduced by down-sampling and reconstructed by up-sampling without significant synthesis errors. For these reasons, depth map contained in 3D video data can be effectively compressed by a combination of spatial resolution reduction and frequency based compression.

Nevertheless, depth map reconstructed by linear down-sampling and up-sampling often generates some synthesis artifacts at the edges of objects. To avoid these artifacts, approaches for up-sampling of reduced resolution depth map have been proposed [7]. For instance,[6] proposes weighted mode filtering that estimates unknown depth values in high resolution by computing a localized histogram of known depth values in low resolution on the basis of color similarity in corresponding original resolution Texture. It is based on the properties of depth map that it has piecewise smoothness but it also has high geometric similarity to Texture. It provides better edge-preserving performance.

However, these reduction and reconstruction processes in recent studies cannot apply rate-distortion control because they work out of the coding loop. The rate-distortion controller for depth map coding regularly optimizes the bit-rate of depth map and the quality of synthesized views. This means that the quality of synthesized views is not always guaranteed with the out-loop resolution reduction of depth map.

2.2. Reduced Resolution Update Mode

RRU is achieved by encoding prediction residuals at reduced resolution, which was obtained by performing prediction using full resolution references. The outline of the basic algorithm is shown in Fig.1. In essence, the residuals need to be down-sampled at the encoder, while transform and quantization of the residuals is performed at lower resolution. For image reconstruction at full resolution, the residuals need to be up-sampled after inverse quantization and inverse transform and the up-sampler needs to be fully specified at both encoder and decoder.



(a) Encoder

(b) Decoder

Fig. 1: Coding scheme with RRU mode

Basic RRU techniques for 2D video compression are achieved by basic linearly interpolated down-sampling and up-sampling. This can effectively reduce the bitrate of a video in RRU mode in cases where prediction error residuals are piecewise smooth. On the other hand, RRU mode is not effective in cases where the residuals are not piecewise smooth. Fig. 2 shows an example of up-sampling error of a residual when applying RRU to the edge of an object on depth map. However , it is not possible to reconstruct residuals with the same up-sampling process as that used for native depth map because down-sampled low resolution residuals lose the properties of native depth map. Therefore, it is difficult to recover lost information on the edges of depth map with the RRU scheme.

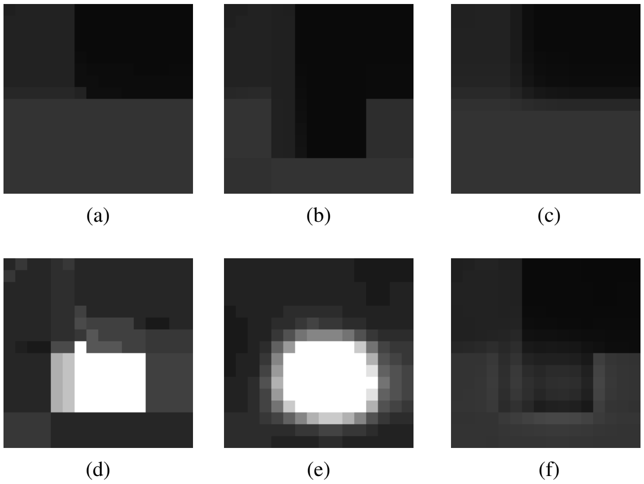


Fig. 2: (2a) Original depth map. (2b) Predicted depth map. (2c) Down-sampled & up-sampled original depth map. (2d) Original residual. (2e) Down-sampled & up-sampled residual on the basis of linear filtering. (2f) Reconstructed depth map with down-sampled & up-sampled residual.

3. PROPOSED METHOD

This paper proposes an advanced RRU mode for depth map compression. The proposed coding scheme is mostly based on the RRU scheme, but differs from it in three ways:

1. Down-sampling is achieved by adaptive sub-sampling depending on Texture.

2. There is a pre-reconstruction process that reconstructs a part of pixels on coded depth map from inverse transformed sub-sampled residuals.

3. Instead of a process for up-sampling residuals, there is a process for up-sampling native depth values. The up-sampling is achieved by adaptive interpolation between pre-reconstructed pixels depending on texture.

Fig. 3 shows the proposed process. It aims to recover unknown native depth values of each non-sampled pixel from reconstructed native depth values on each sampled pixel. Because this process resembles the external re-sampling process of the conventional out-loop re-sampling process, it is easy to apply an effective up-sampling method based on the special properties of native depth map. However, because of a sparsity of prediction residuals, typical linear-sampling methods for resolution reduction such as linear-interpolation or linear-sub-sampling are wasteful. To enable sub-sampled residuals to have as much information as possible, we propose adaptive sub-sampling based on residual energy estimation depending on texture and predicted images of a target depth map.

3.1. Adaptive sub-sampling

To determine the position of the sampled pixel in the original resolution, we model an energy function under the assumption that depth map prediction errors extensively occur around the edges of an object in a target depth map and the edges of an object in its predicted image. The latter can be detected from a predicted image, and the former can be detected from a texture. The energy function is described as follows:

$$E = |\Delta Tex| + |\Delta Prd| \tag{1}$$

In this paper, we divide each transform unit into small blocks as sampling units, and determine a sampled pixel such that it has a maximum energy value on each sampling units. The sampled pixel p_s of the sampling unit U is described as follows:

$$p_s = \arg\max_{q \in U} E(q) \tag{2}$$

Fig. 4 shows the example of sampled pixels and the corresponding reconstructed image.

3.2. Adaptive up-sampling

We also use a weighted mode filter [6] as a non-linear up-sampler. It usually computes the localized histogram of unknown depth values at non-sampled pixels on the basis of known reconstructed depth values at a sampled pixel and color values of texture. In this paper, it additionally considers the depth value of the predicted image of the target depth map. The final reconstructed depth value D of pixel p can be computed as follows:

$$D(p) = \arg\max_{d} H(p, d) \tag{3}$$

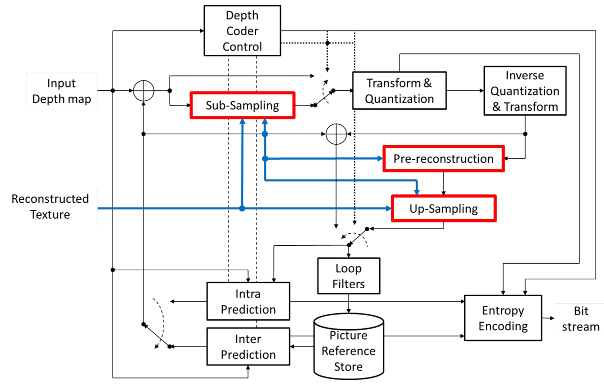
where $H(p, d)$ is the localized histogram around pixel p, computed as follows:

$$H(p, d) = \sum_{q \in N(p)} w(p, q) G_r(d - d_q) \tag{4}$$

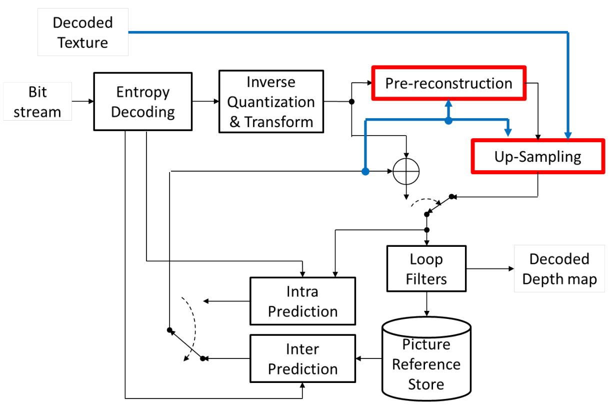
$N(p)$ represents a set of neighboring pixels of the pixel p. G_r is a gaussian function directed by a parameter σ_r and d_q is a pre-reconstructed depth value of pixel q. $w(p, q)$ is the weighting function extendet into the cross-birateral filtering by using the texture and the predicted depth map as guide signals. It can be computed as follows:

$$w(p, q) = G_t(T_p - T_q) G_p(P_p - P_q) G_l(p, q) \tag{5}$$

where G_t, G_p and G_l are Gaussian function directed by a parameter σ_t, σ_p and σ_l. G_t represents color distance measure of corresponding texture T, G_p also represents depth distance measure of predicted depth map P and G_l represents spacial distance measure.



(a) Encoder



(b) Decoder

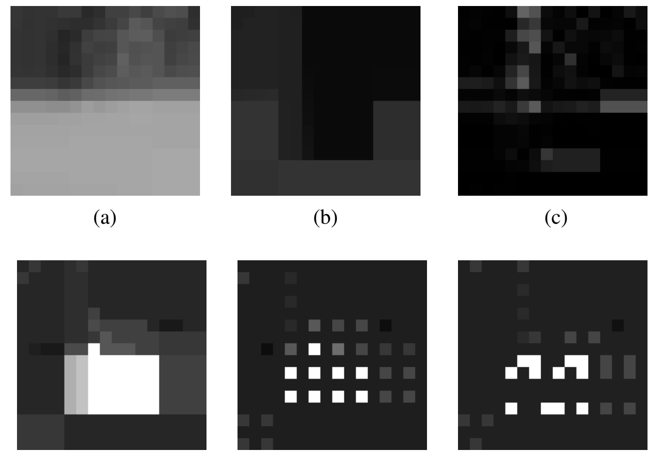
Fig. 3: Additional process of proposal

4. EXPERIMENTAL RESULT

We implemented the proposed method and the RRU scheme on HTM5.1, which is the reference software used in standardization activity for the 3D extension of HEVC [1], and conducted coding experiments to determine its performance. Sampling filters used in RRU are a sinc filter as the down-sampling filter and a lanczos filter as the up-sampling filter. We encoded full frames for each sequence under the common test conditions of HTM5.1 [1] except for QP values for depth map coding. We set the QP values of depth map to be the same as those for texture in every comparison. Weighting parameters σ_r, σ_t, σ_p and σ_l are set to 16, 6, 6 and 3 respectively. Every comparison is based on the average bitrate of coded textures and depth maps, the PSNR of synthesized views.

In the current implementation, both the proposed and RRU schemes are activated only for the TU level. Each mode is activated by a syntax element called an RRU flag. Both methods reduce the resolution of residuals into a quarter. This means the size of a sampling unit in the proposed method is 2x2.

First, as shown in Table 1, we compared the results obtained with the regular RRU and with the proposed method with those obtained with the original HTM. Second, as shown in Table 2, we compared the results obtained using the proposed method with those obtained using the regular RRU. Fig. 5 shows the RD performance between depth coding bitrate and synthesized view distortion obtained for each method.



(a) (b) (c)

(d) (e) (f)

Fig. 4: (4a) Texture. (4b) Predicted depth map. (4c) The intensity of the edge on texture and predicted depth map. (4d) Original residuals. (4e) Sub-sampled residuals on the basis of linear sampling. (4f) Sub-sampled residuals on the basis of linear sampling.

Table 1 shows that both in-loop reduced resolution depth coding methods are effective in reducing bitrate for 3D video compression. Table 2 shows the proposed method was predominant in most cases. Fig. 5 shows the RD performance between depth coding bitrate and synthesized view distortion obtained for each method in both the best and worst cases. In the best case, for the Gt_Fly, the proposed method achieved more effective bitrate reduction results than the the regular RRU. Fig. 5a shows that the regular RRU is effective at the highest bitrate but that its performance decreases at lower bitrates in this case. In the worst case, for Kendo, RRU shows somewhat better performance. However, it is obtained only at the highest bitrate (see Fig.5b). As can be seen, while the improvement obtained depends on the sequence, significant improvement is obtained with the proposed method in most cases.

5. CONCLUSION

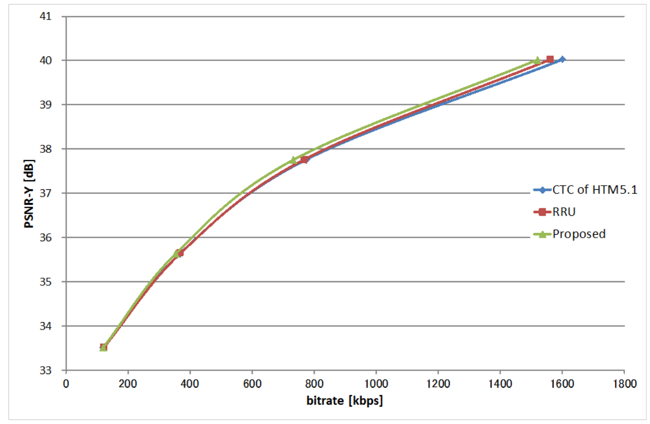
We proposed a novel in-loop reduced resolution depth coding scheme based on the scheme of Reduced Resolution Update (RRU) mode. We also proposed an effective method for reduction of resolution for residuals and reconstruction of depth map on the basis of texture and predicted depth map. The proposed scheme was integrated with HTM, which is the

Sequence	ΔRate [%]		ΔPSNR [dB]	
	RRU	Proposed	RRU	Proposed
Balloons	- 1.42	- 1.50	0.16	0.17
Kendo	- 2.14	- 1.69	0.24	0.18
Newspaper	-1.84	- 2.09	0.19	0.22
GT_Fly	- 0.15	- 0.99	0.02	0.08
Poznan_Hall2	- 0.03	- 0.42	0.00	0.02
Poznan_Street	- 0.46	- 0.96	0.04	0.06
Undo_Dancer	- 0.07	- 0.06	0.00	0.00
XGA	- 1.79	- 1.76	0.20	0.19
Full HD	- 0.18	- 0.61	0.02	0.04
average	- 0.87	- 1.10	0.09	0.11

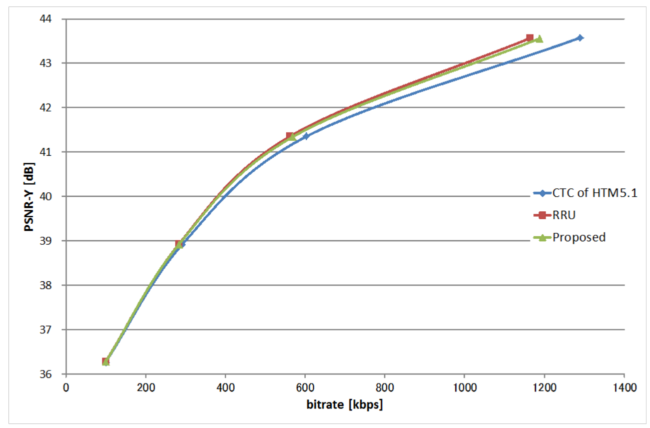
Table 1: HTM5.1 vs RRU and Proposed

sequence	ΔRate [%]	ΔPSNR [dB]
Balloons	- 0.08	0.01
Kendo	0.45	- 0.05
Newspaper	- 0.29	0.03
GT_Fly	- 0.84	0.06
Poznan_Hall2	- 0.40	0.02
Poznan_Street	- 0.51	0.02
Undo_Dance	0.01	0.00
XGA	0.03	- 0.01
Full HD	- 0.43	0.03
average	- 0.24	0.01

Table 2: RRU vs Proposed



(a) GT_Fly



(b) Kendo

Fig. 5: RD performance between depth coding bitrate and synthesized view distortion.

test software of the international standard for 3D extension of HEVC, as an additional residual coding tool, and its performance was evaluated. Experiments showed that it is possible to achieve bitrate reductions of up to 2.09% and about 1.10% on average for seven sequences relative to HTM5.1.

6. REFERENCES

[1] ITU-T SG 16 WP 3 11 and ISO/IEC JTC 1/SC 29/WG. 3D-HEVC Test Model 2. In *JCT3V-B1005_d0*, Shanghai, CN, 2012.

[2] Roger Cheng and Klara Nahrstedt. Empirical study of 3D video source coding for autostereoscopic displays. In *Proceedings of the 15th international conference on Multimedia - MULTIMEDIA '07*, pages 573–576, New York, New York, USA, 2007. ACM Press.

[3] Christoph Fehn and Heinrich-hertz-institut Hhi. Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV. volume 5291, pages 93–104, 2004.

[4] ISO/IEC JTC1/SC29/WG11. Applications and Requirements on 3D Video Coding. In *N11061*, Xian, China, 2009.

[5] ISO/IEC JTC1/SC29/WG11. Call for Proposals on 3D Video Coding Technology. In *N12036*, Geneva, Switzerland, 2011.

[6] Dongbo Min, Jiangbo Lu, and MN Do. Depth Video Enhancement Based on Weighted Mode Filtering. *Image Processing, IEEE Transactions on*, 21(3):1176–1190, 2012.

[7] KJ Oh, Sehoon Yea, Anthony Vetro, and YS Ho. Depth reconstruction filter and down/up sampling for depth coding in 3-D video. *Signal Processing Letters, IEEE*, 16(9):747–750, 2009.

[8] Alexis Michael Tourapis. Reduced Resolution Update Mode for Advanced Video Coding. In *VCEG-V05*, Munich,DE, 2004.