Global Motion Information Based Depth Map Sequence Coding

Fei CHENG[1]*, Jimin XIAO[1], Tammam TILLO[1], and Yao ZHAO[2]

[1]Department of Electrical and Electronic Engineering,
Xian Jiaotong-Liverpool University (XJTLU),
111 Ren Ai Road, SIP, Suzhou, Jiangsu Province,
P.R. China 215123
{first-name.second-name}@xjtlu.edu.cn
http://www.mmtlab.com
[2]Beijing Jiaotong University
Institute of Information Science
Beijing, China
yzhao@bjtu.edu.cn

Abstract. Depth map is currently exploited in 3D video coding and computer vision systems. In this paper, a novel global motion information assisted depth map sequence coding method is proposed. The global motion information of depth camera is synchronously sampled to assist the encoder to improve depth map coding performance. This approach works by down-sampling the frame rate at the encoder side. Then, at the decoder side, each skipped frame is projected from its neighboring depth frames using the camera global motion. Using this technique, the frame rate of depth sequence is down-sampled. Therefore, the coding rate-distortion performance is improved. Finally, the experiment result demonstrates that the proposed method enhances the coding performance in various camera motion conditions and the coding performance gain could be up to 2.04 dB.

Keywords: Depth map, global motion information, down-sampling

1 Introduction

Due to the rapid development of the range and depth sensing technology, depth cameras such as Microsoft Kinect and SwissRange SR4000 [1] have been developed. Depth maps are widely employed in the texture-plus-depth representation for 3D video coding.

A depth map, which represents the distance from the objects in the scene to the capturing camera, together with its aligned texture, have been exploited to describe 3D scenes. Multi-view Video plus Depth (MVD) format is a promising way to represent 3D video content, and recently extensions supporting for the

* This work was supported by National Natural Science Foundation of China (No.61210006, No.60972085).

MVD format have been introduced [2, 3]. With the MVD format, only a small number of texture views associated with their depth views are required. At the decoder or display side, Depth-Image-Based Rendering (DIBR) [4, **?**] is used to synthesize additional viewpoint video.

In the DIBR based 3D video coding scheme, depth map is represented as a gray-scale image, which is encoded independently. A texture and its corresponding depth map describe the features of the same scene in terms of content and distance respectively. The correlation between them should be exploited by an encoder to reduce the redundancy. In [6], the coding performance of depth maps is improved by taking into account the Motion Vector (MV) of texture. This can reduce the time of the Motion Estimation (ME) for depth map encoding due to the reduced coding complexity. Furthermore, it is proposed to add the 3D search to expand the ME of depth map.

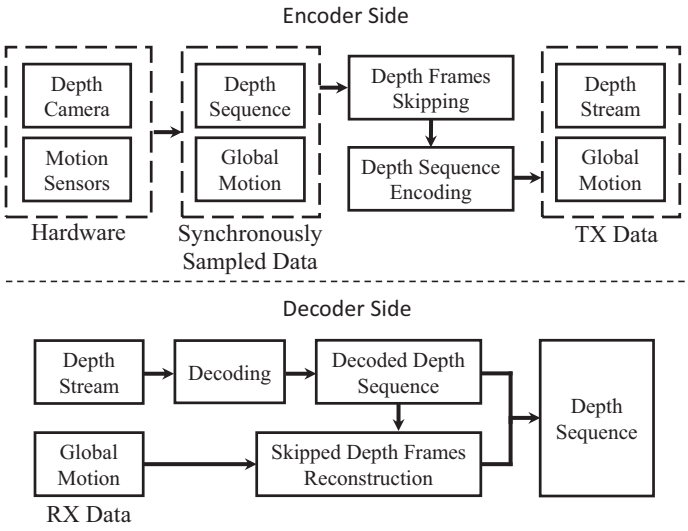

Fig. 1. The diagram of proposed depth map sequence encoding and decoding method

A depth map contains position information of each object, which can be exploited to project the neighboring frames by using the camera global motion information. In this paper, we propose a novel global motion information based depth map sequence coding method. As shown in Fig.1, the synchronously sampled global motion information of a depth camera has been exploited to improve the depth map sequence coding performance. We intentionally skip some depth map frames or blocks during encoding according to the amount of camera global motion. Then, the skipped parts are projected from their neighboring frames using the global motion information between frames. As the bitrate is

reduced, the proposed method achieves the global rate-distortion performance gain.

We develop the hardware prototype to simulate the camera global motion and produce different sequences to test the proposed method under different conditions. In order to simplify the experiment, we test the proposed method under the H.264/AVC standard. The experimental results demonstrate that the proposed method can improve the coding performance compared to H.264/AVC. The average gain could be up to 2.04 dB. It is worth mentioning the proposed method is independent of the coding tools, which means that it can be used with other video coding standards.

The rest of this paper is organized as follows. In Section 2, the details of the proposed method are described. Then, the experimental methods of the proposed scheme and the results are presented in Section 3. Lastly, Section 4 concludes this work.

2 Proposed Method

In many video capturing and depth map sampling scenarios, the camera is moving. The camera global motion leads to the change of image content. As the depth information is represented as gray-scale map after quantization, the gray level changes with the change of depth value. According to the imaging principle, the impact of global camera motion on depth map can be pictorially presented in Fig.2. A cube and a cylinder as examples are captured by a depth camera. The cylinder is farther away from the camera than the cube. The depth map is quantized linearly, while the gray levels (black to white) represent the distance (near to far). The depth camera samples a new depth image after dolly and tracking. With the dolly motion, the depth camera moves forwards. Therefore, the two objects are enlarged with different scales. The scale of the cube is larger than that of the cylinder as it is closer to the camera. At the same time, the gray levels of two objects become darker. For camera tracking, the position of each object is shifted, meanwhile the relative distance between them is also changed. But the gray level of each object dose not change.

The motion information of the depth camera could be obtained using the motion sensor. The depth information together with the relative position can be exploited to project the neighboring frames to the current position. The projected depth map should be similar with the real depth frame.

With the similar principle, some depth map frames or some blocks of one depth frame can be skipped for encoding. Instead, only the global motion information is transmitted to the decoder side. The skipped frames or blocks are reconstructed by projecting from neighboring depth map frames. By reducing the encoding frames or blocks, the total bitrate decreases. Finally, the overall rate-distortion performance is improved.

In summary, two key procedures have to be implemented in order to achieve the proposed depth map sequence coding method. Firstly, some depth map frames or blocks should be skipped for the proposed method. Secondly, at the

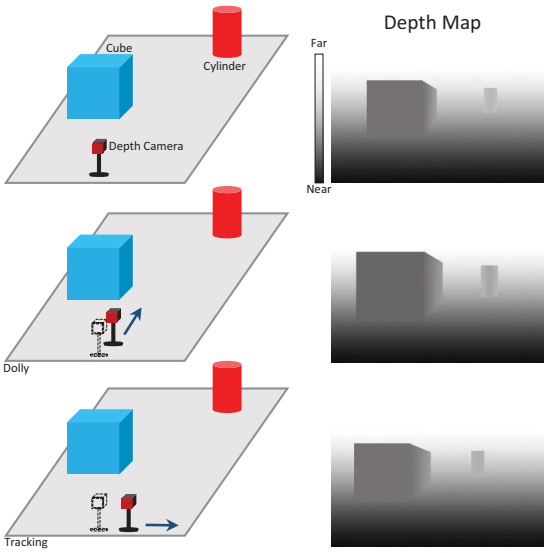

Fig. 2. The impact of camera global motion on depth map

decoder, skipped frames or blocks are projected from the neighboring frames. It is worth mentioning that the transmitted global motion information can also benefit many other applications, such as deblurring and background extraction for moving cameras. The details of each procedure are introduced as follows.

2.1 Depth Map Skipping

Whether the skipped depth frame could be properly projected from the neighboring depth frames is related to two main factors. One factor is the amount of camera global motion. The smaller the motion, the more similar the current frame to the neighboring frames, which means it is not difficult for the decoder to project the skipped depth frames. Therefore, the depth frame skipping can be decided dynamically by the amount of global motion. The threshold of skipping is evaluated based on the content of the sequence.

Another factor is the change of contents in the scene. The moving objects are difficult to be projected from their neighboring frames. Therefore, blocks containing moving objects are segmented and encoded separately, whereas, the rest blocks of the current frames are skipped. It is worth mentioning that both sides of the neighboring frames should be projected to the current position. The decision on whether the frames or blocks are skipped or not is based on the differences between the current frame and the projected frames (blocks). Large difference means that the frame (block) is skipped, and vice versa.

An example of depth map frame skipping is described in Fig.3. The depth frames D_{n+1}, D_{n+3} and D_{n+4} are skipped at the encoder side. Then, D_{n+1} is projected from D_n and D_{n+2}, while D_{n+3} and D_{n+4} are projected from D_{n+2}

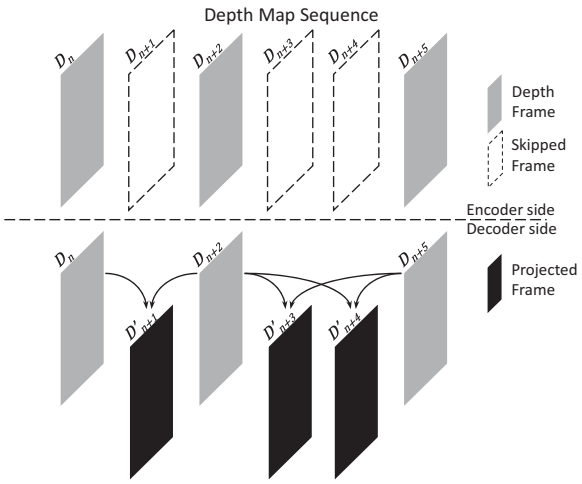

Fig. 3. An example of depth map frame skipping and projection

and D_{n+5}. Finally, each frame of the depth sequence is reconstructed at the decoder side.

2.2 Depth Map Projection

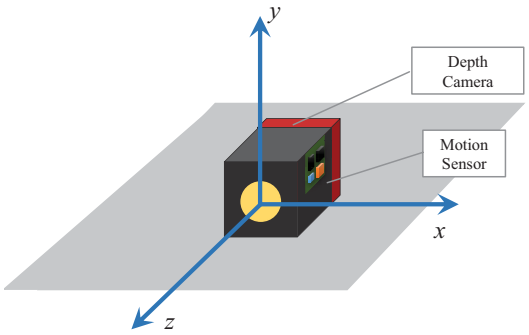

Fig. 4. The 3D coordinate system definition for the camera

Each pixel in the depth map needs to be converted from a 2D point into 3D coordinate space for 3D projection. First, the 3D coordinate system needs to be defined as shown in Fig.4. The x-axis and the y-axis are parallel to the image, while the z-axis represents the depth.

Let $\mathbf{P} = [w, h]$, where w and h represent the horizontal and vertical coordinate of a pixel in the image. As the depth z of each pixel has been quantized to

an integer n in the depth map frame, then it needs to be dequantized by:

$$z = Q^{-1}(n), \tag{1}$$

where Q is the quantization method of depth map sequence. The 3D homogeneous coordinate of a pixel can be converted by:

$$\mathbf{C} = [x, y, z, 1] = \left[K(\frac{W}{2} - w) \cdot z, K(\frac{H}{2} - h) \cdot z, z, 1 \right], \tag{2}$$

where W and H are horizontal and vertical resolution of the depth map respectively. K is the intrinsic parameters of the depth camera, which is represented as:

$$K = \frac{\tan(\phi_w)}{W} = \frac{\tan(\phi_h)}{H}, \tag{3}$$

where ϕ_w and ϕ_h are the horizontal and vertical angles of the view respectively. The 4×4 projective transformation matrix is represented by $\mathbf{T}$, which is related to the translation and rotation from the neighboring position to the current position. The new coordinate of a pixel on the project frame can be obtained by:

$$\mathbf{C}_p = \mathbf{C} \times \mathbf{T} = [x_p, y_p, z_p, 1]. \tag{4}$$

The 3D coordinate of each pixel in the current position has to be inversely converted to 2D coordinate in the depth map:

$$\mathbf{P}_p = [w_p, h_p] = \left[\frac{W}{2} - \frac{x_p}{K z_p}, \frac{H}{2} - \frac{y_p}{K z_p} \right]. \tag{5}$$

As the global motion might change the depth of each pixel, n' of each pixel needs to be quantized to form a reconstructed depth frame from the projected depth information by using the same quantization method Q:

$$n' = Q(z_p). \tag{6}$$

Finally, each pixel of the neighboring frames can be transformed to a new 2D coordinate in the depth frame of the current position. However, some of them might be located outside of the image and some of them might not be integers, which leads to holes. Therefore, an interpolation algorithm is utilized to fill holes and smooth the projected depth frame, which is represented as $\mathbf{D}'$.

3 Experimental method and Results

To the best of our knowledge, there is no standard sequence where the proper (not estimated) global motion information is available. Consequently, we produced some sequences with synchronously sampled global motion information using our platforms. We tested the proposed scheme using the sequences we produced. These sequences are available for download at http://mmtlab.com/dmcmb.

3.1 Data Acquisition and Prototypes

In this paper, we developed a programmable track slider as shown in Fig.5 (the texture camera is not used in this paper). A shaft encoder is employed as the motion sensor to get the accurate translational distance. The depth camera is a Mesa Imaging SwissRanger SR4000.

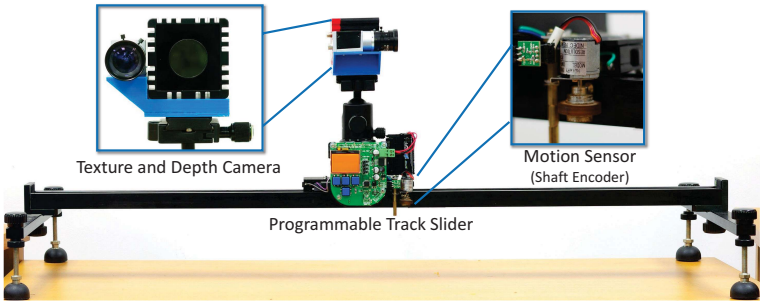

Fig. 5. The customized prototype for the translational motion

3.2 Experiments and Results

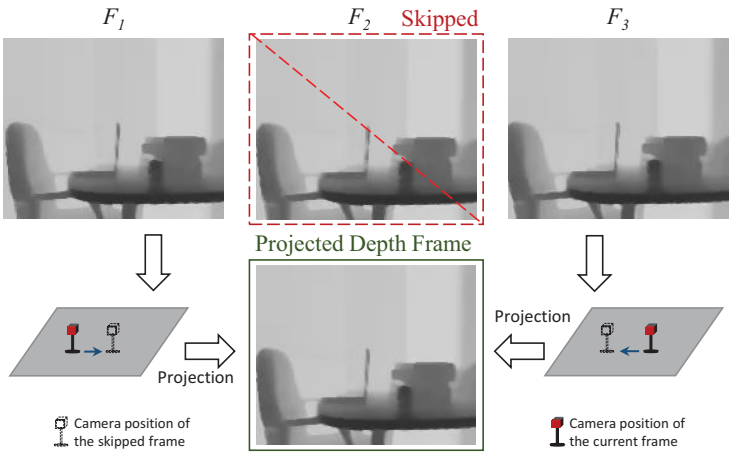

Fig. 6. An example of the processing of the tracking right motion experiment

In the experiment, we used forward movement (dolly) and the right movement (tracking) as examples to test the proposed method. The resolution of depth map

generated by SR4000 is QCIF (176×144) @ 25 fps. The depth map quantization method is uniform, where white presents farthest distance and black presents nearest distance. To evaluate the proposed method, H.264/AVC JM reference software 14.1 [7] is used. The projection program is developed based on the principle mentioned in Section 2. To span a reasonable range of bitrate, the QP was differently set for the proposed method and the standard H.264/AVC. As the frame skipping scheme in the proposed method decreases the bitrate, we have to adjust QP values in order to obtain similar range of bitrate. In the forward dolly experiment, the QP value for the proposed method test is set from 24 to 50, while for standard H.264/AVC it is set from 26 to 50. In the right tracking experiment, the QP value for the proposed method test is set from 26 to 50, while for standard H.264/AVC it is set from 28 to 50. Fig.6 illustrates an example of the right tracking experiment.

Fig.7 presents the PSNR comparison between the proposed method and standard H.264/AVC. The BD-Rate is -29.7%, while the BD-PSNR is 1.29 dB. In Fig.8, the PSNR comparison between the proposed method and standard H.264/AVC is presented. The BD-Rate is -41.12%, while the BD-PSNR is 2.04 dB.

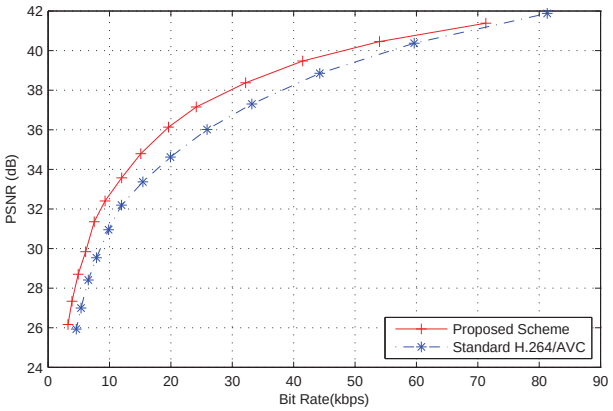

Fig. 7. PSNR versus bitrate for the proposed scheme and standard H.264/AVC in dolly motion; skipping one frame from two frames

From the results, we could conclude that the gain of the tracking motion is larger than that of the dolly motion. The reason is that in the dolly motion, each object of the depth map is scaled, and the interpolation needs to be employed. This would reduce the accuracy of the reconstructed frames.

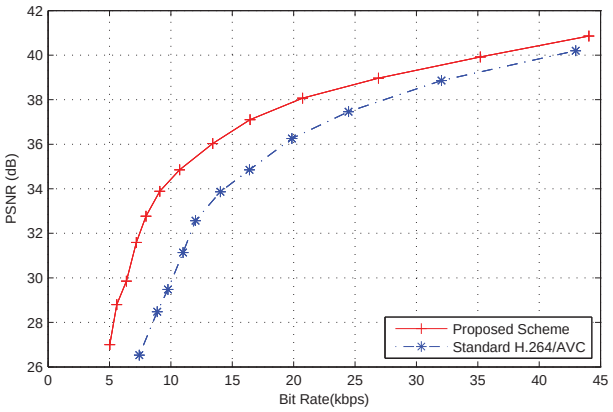

Fig. 8. PSNR versus bitrate for the proposed scheme and standard H.264/AVC in tracking motion; skipping one frame from two frames

4 Conclusions

This paper has introduced a novel depth map sequence coding method using the camera motion information. Compared with the existing H.264/AVC standard, the proposed scheme is able to improve the coding performance up to 2.04 dB. It is noticed that the accuracy of the depth and motion information affects the performance of the proposed method. In the future, we will improve the data precision and test more types of motion, such as rotation and combined motion.

References

1. MESA IMAGING, "SR4000," http://www.mesa-imaging.ch/products/sr4000/, [Online].
2. M Hannuksela, Y Chen, T annd J.-R. Ohm Suzuki, and G. Sullivan (ed.), "Avc draft text 8," *JCT-3V document JCT3V-F1002*, vol. 16, 2013.
3. Ying Chen, Miska M Hannuksela, Teruhiko Suzuki, and Shinobu Hattori, "Overview of the mvc+ d 3d video coding standard," *Journal of Visual Communication and Image Representation*, vol. 25, no. 4, pp. 679–688, 2014.
4. Philipp Merkle, Aljoscha Smolic, Karsten Muller, and Thomas Wiegand, "Multi-view video plus depth representation and coding," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. IEEE, 2007, vol. 1, pp. I–201.
5. Christoph Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 93–104.
6. Pei-Jun Lee and Xu-Xian Huang, "3d motion estimation algorithm in 3d video coding," in *System Science and Engineering (ICSSE), 2011 International Conference on*, June 2011, pp. 338–341.
7. HHI Fraunhofer Institute, "H.264/AVC Reference Software," http://iphome.hhi.de/suehring/tml/download/, [Online].