Signal Processing: Image Communication 47 (2016) 183–192

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Fast depth map coding based on virtual view quality

Chao Yang [a,b], Ping An [a,b,*], Liquan Shen [a,b,*], Ran Ma [a,b], Nina Feng [a,b]

[a] *Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai, China*
[b] *School of Communication and Information Engineering, Shanghai University, Shanghai, China*

ARTICLE INFO

Article history:
Received 21 January 2016
Received in revised form
24 May 2016
Accepted 25 May 2016
Available online 3 June 2016

Keywords:
Depth map
Virtual view distortion
Rate distortion optimization
Quadtree structure
3D-HEVC

ABSTRACT

Multi-view video plus depth (MVD) format is a three-dimensional (3D) video representation. The depth map in MVD provides the scene distance information and is used to synthesize the virtual view with texture image through Depth Image Based Rendering (DIBR) technique. In this paper, we first present a virtual view distortion estimation (VVDE) scheme to estimate the synthesized virtual view distortion induced by depth map compression. Then based on the synthesized virtual view distortion, we propose region based Quantization Parameter (QP) adjustment scheme to improve the perceived virtual view quality, fast VVDE scheme in depth map coding Rate Distortion Optimization (RDO) to improve the coding efficiency and fast Coding Unit (CU) partition scheme to reduce the depth map coding complexity. Experimental results demonstrate that the proposed algorithm can achieve 0.50 dB gain with 47.37% coding time reduction.

1. Introduction

THREE-DIMENSIONAL (3D) video has drawn a lot of attention since 3D movies won great success in the market. Multi-view video plus depth (MVD) 3D video format enables functionalities like 3D television [1] and free viewpoint video [2]. MVD consists of multiple views of colorful texture image and gray depth map. The depth map contains scene geometric information and is used to synthesize the virtual view through Depth Image Based Rendering (DIBR) technique [3].

The High Efficiency Video Coding (HEVC) standard [4] was amended to 3D-HEVC [5] for MVD coding. In 3D-HEVC, texture image and depth map are jointly coded and put in the same bit stream, and depth map is used to provide geometric information to synthesize the intermediate virtual views. The bitrate of the depth map accounts for about 40–60% of the texture image bitrate [6], which makes MVD possible to reduce the total bitrate of 3D video.

Depth map has some different features compared with texture image. Depth map is a gray-scale image containing the geometric information, and every pixel in texture image has its corresponding pixel in depth map in MVD.

Depth map coding error induces geometric information error which will cause distortion in the synthesized virtual views and affect the 3D viewing experience. To estimate the virtual view distortion induced by depth map compression, various methods have been proposed. [7–9] proposed a virtual view distortion model to allocate bit between texture image and depth map, and [10–12] proposed a linear distortion model to estimate virtual view distortion induced by depth map compression and use the model to improve depth map coding efficiency. De Silva et al. proposed an algorithm to minimize the virtual view distortion in depth map intra mode coding [13,14], but the algorithm is of great time consuming. Kim et al. proposed rendered view distortion estimation model with new Lagrangian Multiplier (LM) to improve depth map coding efficiency [15]. In 3D-HEVC Test Model (HTM) reference software, Synthesized View Distortion Change (SVDC) [16] and View Synthesis Distortion (VSD) are performed jointly as View Synthesis Optimization (VSO) to improve the depth map coding efficiency [5]. SVDC directly renders the virtual views by using the decoded texture image and depth map at the encoding side, which has a high complexity. VSD is a model based virtual view distortion scheme, which has lower complexity compared with SVDC.

The high coding complexity is a key factor that limits the 3D-HEVC in practical applications. Therefore it is necessary to develop fast coding methods to reduce the high coding complexity. Quadtree-based coding structure [17] is used in HEVC to improve the coding efficiency. In the quadtree-based coding structure, a frame is divided into several Coding Tree Units (CTUs), the CTUs can then be split into four smaller coding units (CUs), and a specified maximum depth level is set to limit the CU split recursion.

* Corresponding authors at: Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai, China.
E-mail addresses: anping@shu.edu.cn (P. An), jsslq@shu.edu.cn (L. Shen).

http://dx.doi.org/10.1016/j.image.2016.05.015
0923-5965/

The depth map needs to be coded in different CU sizes and based on the Rate Distortion Cost (RDC) to decide the best CU size. This Quadtree structure can improve the depth map coding efficiency but has extremely high complexity because of the recursion process. The implementation of the VSO technique can improve the synthesized virtual view quality but will also result in extra coding complexity.

Low complexity color video coding has been widely studied, and fast CU partition algorithms have been widely exploited. Shen et al. proposed an algorithm to determine the CU size based on the sizes of spatially and temporally neighboring CUs to reduce the video encoding complexity [18,19]. Zhang et al. proposed a fast CU size decision algorithm based on the characteristic of color video [20]. Fast Intra-mode decision algorithm using Inter-level and spatiotemporal correlations was proposed in [21]. For 3D video encoding, Kuo et al. proposed fast CU partition, fast prediction unit (PU) partition and search range reduction algorithms by referring to the depth edge information [22].

Some other fast depth map coding algorithms have also been proposed to reduce the depth map coding complexity. Shen et al. proposed low complexity depth map coding algorithm by incorporating fast mode decision and motion estimation strategies [23]. Ma et al. proposed a low complexity view synthesis optimization scheme in 3D video coding by devising a novel zero-synthesized view difference (ZSVD) model [24]. Statistics based fast CU depth decision algorithm is proposed in [25]. However, these algorithms considered only the characteristic of depth map, not the quality of the synthesized virtual view.

In our previous work [26] prior to this research, we proposed a virtual view distortion estimation (VVDE) model based on the characteristics of both texture image and depth map and implement the model in the Rate Distortion Optimization (RDO) process to improve depth map coding efficiency.

In this paper, we employ the VVDE model and simplify the RDO process and adjust Quantization Parameters (QPs) in different depth map regions based on the virtual view distortion to improve the depth map coding efficiency with low complexity. And a fast CU partition scheme based on the virtual view distortion is proposed to further reduce the depth map coding complexity.

The rest of this paper is organized as follows. Section 2 introduces the VVDE model which estimate the virtual view distortion induced by depth map compression and QP adjustment scheme based on the virtual view distortion. Section 3 introduces fast depth map coding algorithm based on virtual view distortion, and the overall algorithm is presented in Section 4. The experimental results and conclusions are given in Section 5 and Section 6, respectively.

2. Virtual view distortion optimization

Depth map is used as a sort of supplemental data for virtual view synthesis rather than an independent video data for display in 3D video system. It is important to take the synthesized virtual view quality into consideration in depth map encoding. Therefore, we propose a distortion model to estimate the virtual view distortion induced by depth map coding and improve the depth map coding efficiency based on the model.

2.1. VVDE scheme

Depth map is a gray image recording the distance information of objects from capturing devices [27]. Different from texture image, depth map rarely contains any texture and is predominantly flat with sharp edges making boundaries between objects at different depth [28].

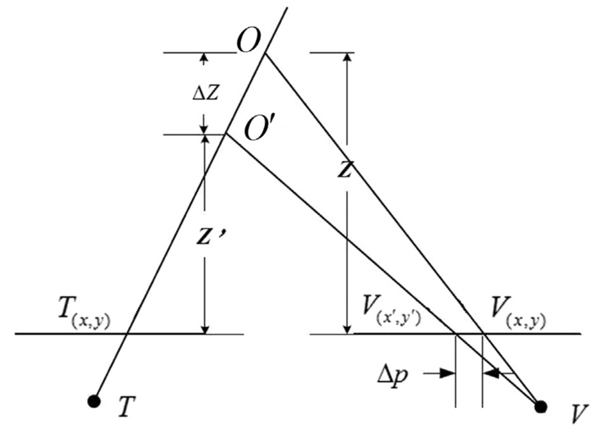

Fig. 1. DIBR algorithm illustration, when depth map is uncompressed, $T_{(x,y)}$ in the reference texture image is projected to $V_{(x,y)}$; when depth map is compressed, $T_{(x,y)}$ is projected to $V_{(x',y')}$.

Sum of squared error (SSE) is widely used as a distortion metric in video coding due to its low complexity and good performance. For this reason, we use SSE to calculate the distortion between the ground truth and distorted virtual view as:

$$SSE_v = \sum_{x,y \in (W,H)} \left(V_{(x,y)} - \hat{V}_{(x',y')} \right)^2 \tag{1}$$

where $V_{(x,y)}$ and $\hat{V}_{(x',y')}$ denote the pixel intensity in the ground truth and the distorted virtual view images located at position (x, y) and (x', y'), respectively. W and H are the width and height of the current virtual view image block or frame.

The depth map distortion induces disparity distortion in virtual view synthesis process, as illustrated in Fig. 1. And there exist a linear relationship between the depth map distortion $|\Delta z|$ and disparity distortion $|\Delta p|$ as:

$$|\Delta p| = \left| \frac{f \cdot l}{Z} - \frac{f \cdot l}{Z'} \right| = \frac{f \cdot l}{255} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) \cdot |\Delta z| = \alpha \cdot |\Delta z| \tag{2}$$

where f is the camera focal length and l is the baseline distance between the reference view and the synthesized virtual view, and Z_{near} and Z_{far} denote the nearest and farthest depth values, respectively.

Based on Eqs. (1) and (2), the virtual view distortion induced by depth map distortion can be written as:

$$SSE_v = \sum_{x,y \in (W,H)} \left(V_{(x,y)} - V_{(x+\Delta p,y)} \right)^2 \tag{3}$$

It could be concluded that the virtual view distortion is related to the texture image pixel intensity $V_{(x,y)}$, and the disparity error Δp.

Depth map distortion induces disparity error, which leads to pixel location deviation in the synthesized virtual view. The deviated pixels will be interpolated into the integer grid. For this reason, we consider the distortion of interpolated pixels in texture image between pixels $I_{(x,y)}$, $I_{(x-1,y)}$ and $I_{(x+1,y)}$, and introduce a gradient metric G to evaluate the image complexity written as follows:

$$G = E\left[\sum_{x=2}^{W-1} \sum_{y=1}^{H} \left(I_{(x,y)} - I_{(x-1,y)} \right)^2 + \left(I_{(x,y)} - I_{(x+1,y)} \right)^2 \right] \tag{4}$$

where $I_{(x,y)}$ is the pixel intensity in the image, and $E[\ ]$ denotes the expectation taken over all pixels in one image.