Journal of Visual Communication and Image Representation 54 (2018) 145–154

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

A novel distortion criterion of rate-distortion optimization for depth map coding☆

Ziqi Zheng[a], Junyan Huo[a,*], Bingbing Li[a], Hui Yuan[b], Weisi Lin[c]

[a] State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China
[b] School of Information Science and Engineering, Shandong University, Ji'nan, China
[c] School of Computer Science and Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Keyword:
Three dimensional video (3DV)
Multiview video plus depth (MVD)
Rate-distortion optimization
Virtual view distortion
3D-HEVC

ABSTRACT

In 3D video coding systems, depth maps are not displayed to the viewers, but provide the geometric information to generate virtual views. To ensure the quality of virtual views, the rate-distortion optimization (RDO) in depth map coding adopts the virtual view distortion as the distortion item. The virtual view distortion comes from the reconstructed color video distortion and depth distortion. It is usually recognized that the virtual view distortion caused by reconstructed color video distortion is independent of that in depth map coding. Preliminary experiments reveal that the virtual view distortion in depth map coding is also influenced by the reconstructed color video distortion. Therefore, we proposed a novel distortion criterion of depth map coding in which the reconstructed color video distortion is modeled and joins into the virtual view distortion calculation. Correspondingly, the associated Lagrange multiplier is also proposed. Experimental results demonstrate that the method by integrating the proposed distortion criterion into RDO for depth map coding can achieve an average 12.72% bitrate saving compared with SSD based RDO method and can also lead a bitrate reduction (0.64%) compared with the existing distortion estimation method in the current 3D-HEVC reference software. With the associated Lagrange multiplier, the proposed distortion criterion can achieve 12.98% bitrate saving compared with SSD based RDO method on average.

1. Introduction

Rapid increase in three dimensional (3D) video applications has taken place, such as 3D movie, 3D television (3DTV) and free viewpoint television (FTV) [1]. These applications can provide favorable immersive vision, real depth perception and friendly interactivity. 3D video offers more selections of viewing angles and a better visual experience to viewers. In order to enjoy 3D scene freely, extensive 3D information of a real world should be utilized, but it leads to a huge amount of data [2,3]. Therefore, how to ensure high quality of free view videos and reduce transmission costs become the main challenge of 3D video applications. Multiview video plus depth (MVD) [4] format is one of the most popular 3D video representations, in which color videos and associated depth maps of several selected views are involved. Virtual views in viewer-on-demand positions can be synthesized by using depth image based rendering (DIBR) technology [5,6].

In the MVD representation, multiview color videos and multiview depth maps provide rich information of the 3D scene. Specifically, color videos capture the textural and color information of real world, while depth maps supply the geometric information. Here, both multiview color videos and multiview depth maps are coded and transmitted to the receiver side. A straightforward method to code multiview depth maps is to use the existing coding technologies for color videos. However, due to the different characteristics between color videos and depth maps, these technologies cannot compress depth maps efficiently. Hence, many approaches have been proposed by exploiting the depth map specific characteristics for efficient depth map coding [7,8]. Song et al. [9] proposed a unified depth intra coding method that incorporates depth intra modes into the regular intra mode set. Daribo et al. [10] proposed arbitrarily shaped sub-block motion prediction and arithmetic edge coding for depth map coding based on the observation that neighboring pixels of similar depth have similar motion. Chen et al. [11] proposed a method for virtual viewpoint quality assessment based on the visual masking effect by analyzing the distortion caused by inaccurate depth map and rendering algorithm.

Besides those depth feature based coding algorithms, extensive researches also focus on the rate-distortion optimization (RDO) of depth map coding. The conventional RDO uses the sum of squared differences

☆ This paper has been recommended for acceptance by Zhenzhong Chen.

* Corresponding author.
E-mail address: jyhuo@mail.xidian.edu.cn (J. Huo).

https://doi.org/10.1016/j.jvcir.2018.05.010
Received 6 February 2017; Received in revised form 15 March 2018; Accepted 11 May 2018
Available online 14 May 2018
1047-3203/

(SSD) of depth maps as the distortion for depth map coding. However, considering that depth maps are not displayed to the viewers, but indirectly influence the quality of the virtual views. A better distortion criterion of RDO for depth map coding is designed to adopt the virtual view distortion (D_{vs}) as the distortion item of RDO. The methods in [12,13] aimed at the calculation of the overall virtual view distortion change (SVDC), which is in relation to the change of a depth block from original to distorted depth maps. These studies can calculate the block-level distortion of virtual views exactly as well as involve a high computational complexity.

In order to reduce the computational complexity as well as preserve the rate distortion performance, many algorithms have been proposed. Dou et al. [14] presented a scheme aimed at reducing coding complexity of the SVDC calculating by skipping line segments of pixels with variable lengths based on information from both of color videos and depth maps. Oh et al. [15,16] introduced a virtual view distortion estimation algorithm by analyzing the relationship between depth distortion and D_{vs}. A similar distortion estimation model for virtual views was also proposed in [17] to estimate the pixel-wise distortion of virtual views due to the depth distortion. Yuan et al. [18] studied the disparate geometric errors caused by depth distortion in rendering process and derived a polynomial model for D_{vs}. Li et al. [19] analyzed the rendering process of pixels in different positions and introduced an adaptive model for different pixel intervals to calculate the distortion of virtual views. A ground-truth virtual view PSNR estimation method was put forward in [20]. In addition, Fang et al. [21] took the virtual view position into account and proposed an analytical model to estimate the depth-distortion-induced D_{vs}.

Except the depth distortion, it has been found that D_{vs} is also influenced by the textural characteristics of corresponding original color videos [16,22–24], named as *textural characteristics* in our paper. The textural characteristics in [16] was measured by the spatial gradient of the reconstructed color video. A region based D_{vs} model over different textural characteristics (including texture area and smooth area) was introduced in [22]. Kim et al. [23] used the cross correlation between the local original color video and the shifted original color video to estimate the pixel value in the virtual view and summarized a D_{vs} estimation method. Shao et al. [24] used the texture boundaries of color videos to measure the depth sensitivity, and then proposed a low-complexity depth map coding scheme by incorporating the depth-sensitivity-fidelity characteristic into the RDO process.

It should be noted that virtual views are derived by reconstructed color videos. The textural characteristic of the reconstructed color video is influenced by the lossy compression. For example, a reconstructed color video coded with a large quantization parameter (QP) is more smooth than that coded with a small QP. Motivated by this feature, the impact of the reconstructed color video distortion on D_{vs} is discussed and employed into a novel distortion criterion of RDO process for depth map coding.

In this paper, the relationship among the D_{vs}, depth distortion, textural characteristics and the reconstructed color video distortion is first analyzed in Section 2. Then, based on the analysis, we propose a novel distortion criterion of RDO for depth map coding. The associated Lagrange multiplier is also proposed in Section 3. Experiment results and analyses are provided in Section 4, followed by conclusions in Section 5.

2. Analyses of effect factors for D_{vs}

In a DIBR system [5,6], multiview depth maps are used to provide geometric information for rendering virtual views. Due to the depth map distortion, the geometric information may lead to wrong pixel positions in the rendering process, i.e. *the geometric error*. This error will consequently degrade the quality of virtual views.

For simplicity, a typical horizontally-arranged camera array is discussed in our paper. Through the mutual transformation among the image coordinate, camera coordinate and 3D world coordinate, we can obtain the disparity P [23] by

$$P = cfld, \tag{1}$$

where P represents the pixel offset between an original color video and a virtual view; $c = (1/Z_{\text{near}} - 1/Z_{\text{far}})/255$, Z_{near} and Z_{far} represent the nearest and farthest distances between the scene and the camera, respectively; f represents the focal length; l represents the camera baseline distance; d represents the depth value of a pixel.

As for the geometric error ΔP, it can be expressed as

$$\Delta P = cfl\Delta d, \tag{2}$$

where Δd represents the depth distortion. In the horizontally-arranged camera array case, only horizontal offset of ΔP is considered.

The relationship between the geometric error ΔP and D_{vs} in depth map coding has been discussed in [22]. The conclusion includes: the geometric error ΔP is linear to D_{vs}; the slope of the linear model is relevant to the textural characteristic [22,23]. That is to say, the same geometric error ΔP may lead to different image quality degradations over different texture regions [24].

To investigate the impact of the geometric error ΔP, textural characteristics and the reconstructed color video distortion on D_{vs}, a preliminary experiment is conducted in this section. Similar with [22], we suppose the geometric error ΔP caused by depth distortion is fixed. Different ΔPs that belong to {1, 2, 3, 4, 5, 6, 7} are designed in our experiment. Four typical MVD sequences: *Balloons* [25], *Newspaper* [26], *GhostTownFly* [27] and *UndoDancer* [27], are tested.

Fig. 1 gives a comparison of the Mean Squared Error (MSE) of virtual views over different test sequences with original color videos and different ΔP. The x-axis represents the geometric error ΔP and the y-axis represents MSE of D_{vs}, denoted as MSE_{vs}. It can be observed that MSE_{vs} increased with the geometric error ΔP for every curve. That is to say, the larger the depth distortion Δd, the larger the geometric error ΔP will be, and this will increase the corresponding MSE_{vs}. It means that MSE_{vs} is in proportional to the depth distortion.

For different test sequences, it can also be seen that with the same geometric error ΔP, different MSE_{vs} are derived. This implies that the MSE_{vs} is influenced by the color video content. Specifically, if the geometric error ΔP is small, the area contains rich complex textures will lead to a large D_{vs}, while the mostly smooth local area [23] will lead to an inappreciable D_{vs}. That is to say, textural characteristics have effect on the MSE_{vs}.

To obtain different reconstructed color video distortions, the QP of color videos (denoted as Q_c) are set to 15, 25, 40 and 50. Fig. 2 illustrates the curve of MSE_{vs} under different geometric error ΔP and Q_c. The x-axis and the y-axis are the same with Fig. 1. For better

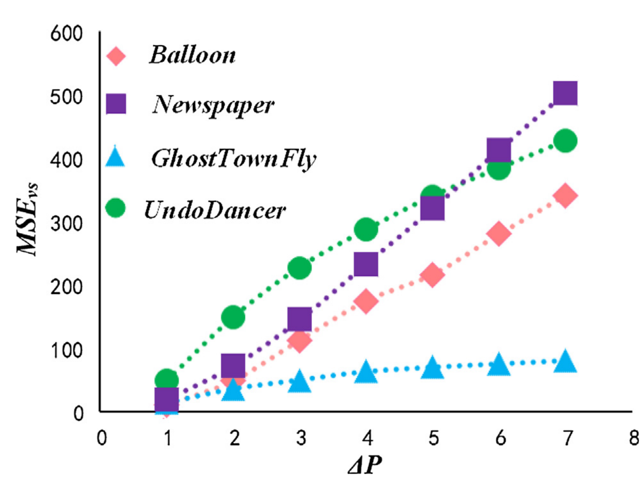

Fig. 1. Comparison of the Mean Squared Error (MSE) of virtual views over different test sequences with zero D_{color} (original color video) and different ΔP.