Signal Processing: *Image Communication* 29 (2014) 962–970

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

An adaptive quantization algorithm without side information for depth map coding

Byung Tae Oh

Korea Aerospace University, Republic of Korea

ARTICLE INFO

Article history:
Received 14 December 2013
Received in revised form
13 June 2014
Accepted 13 June 2014
Available online 25 June 2014

Keywords:
Depth map coding
Compression
3D video
Multiview plus depth map
Quantization parameter
Rate-distortion model

ABSTRACT

This paper presents a novel block-adaptive quantization scheme for efficient bit allocation without side information in depth map coding. Since the type of distortion in a depth map causes different effects in terms of the visual artifacts in a synthesized view, the proposed method adaptively assigns the number of bits according to the characteristics of the corresponding texture block. I have studied the details of the depth map and its rendered view distortion, modeled these analytically, and then proposed a new rate and distortion model for depth map coding. Finally, I derived a simple closed-form solution based on my proposed rate and distortion model, which determines the block-adaptive quantization parameter without any side information. Experimental results show that the proposed scheme can achieve coding gains of more than 0.6% and 1.4% for quarter- and full-resolution depth maps, respectively, in a multi-view-plus-depth 3D system.

1. Introduction

A free view-based three-dimensional (3D) video system is one of a few attractive approaches available these days for a next generation 3D system. Such a system can display a given scene from any angle, so that users experience more realistic 3D effects than those provided by a stereoscopic display system. However, designing a 3D system is a difficult task, mainly because it involves a large amount of multi-view source data, which requires a high transmission bandwidth and a sizeable storage space. One promising solution is the multi-view plus depth (MVD) format that uses the depth image-based rendering (DIBR) technique, as shown in Fig. 1 [1–4]. In this scenario, numerous intermediate views are not processed or transmitted, but are synthesized anew by the decoder instead. This reduces the size of the input data. The Video Coding Experts Group (VCEG) of the International Telecommunications Unions Telecommunication Standardization Sector (ITU-T) and the Moving Picture Experts Group (MPEG) of the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC) have partnered to form the ITU-T/ISO/IEC Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V). Together, these researchers created the H.264/AVC (Advanced Video Coding) format [5]. They also recently developed a high efficiency video coding (HEVC) format [6], and plan to complete JCT-3V standardization by 2015 [7].

Two major approaches have thus far been researched to improve the efficiency of 3D video coding. The first seeks to improve inter-view prediction with the help of a depth map, and the second looks to improve depth map coding. Since a depth map provides completely new content – from capturing to displaying – compared to conventional texture scenes, a number of intensive efforts have been made in the last few years to improve coding efficiency. For example, different types of intra prediction schemes have been introduced in [8,9], as have depth-oriented schemes for in-loop filters [10]. These have provided view

E-mail address: byungoh@kau.ac.kr

http://dx.doi.org/10.1016/j.image.2014.06.004
0923-5965/

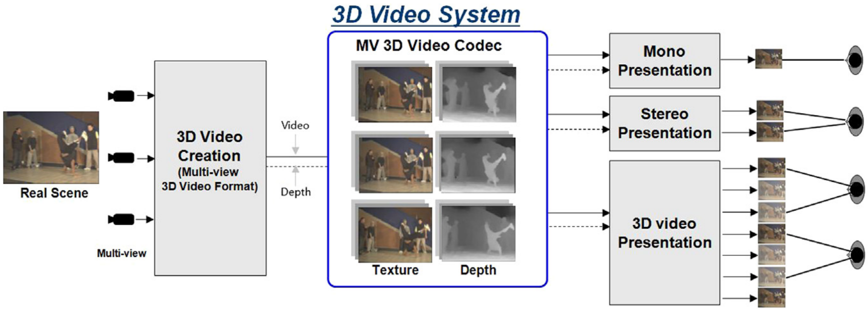

Fig. 1. Overview of the 3DV system.

distortion-based rate distortion (R-D) synthesis optimization tools (VSO) [11,12], which have been adopted in the reference software for 3D extensions of AVC and HEVC. Some of these efforts have been extensively discussed with respect to standardization activities.

In this paper, I propose a new block-adaptive quantization scheme for depth map coding. For this, I first analyze how the distortions in depth maps degrade virtually synthesized views, and then introduce an efficient power model to approximate the relationship in a simple form. With the proposed model, I modify the well-known R-D model and derive a closed-form solution for the determination of the block-wise quantization parameter (QP) without coding or transmitting additional information. Consequently, my model sustains the overall bit-rate while improving the quality of virtual views. To the best of my knowledge, no research has thus far been conducted to adaptively change the block QP according to its contents. Instead, a few conventional approaches have been used to vary the block QP values, mostly targeting rate control as a method of preventing buffer overflows and underflows [13–16]. However, the motivation and details of these approaches are completely different from those of the proposed approach, in that these approaches directly determine the optimal QP for each block and analyze its corresponding texture view.

The rest of this paper is organized as follows. I briefly summarize the depth map and its coding properties in Section 2. I derive my proposed block-adaptive QP determination method in Section 3. In Section 4, I provide experimental results to show the effectiveness of the proposed scheme with respect to coding performance and complexity. Finally, I make concluding remarks in Section 5.

2. Analysis of depth map coding and view synthesis

2.1. Virtual view warping with distortion in depth maps

Since the MVD-formatted 3D video system strongly relies on the DIBR technique, it is important to understand how view synthesis is performed given the depth maps and the camera parameter information. It is possible to generate any view from the arbitrary view position, except holes in occluded regions, if the given depth maps and the camera parameter information are perfectly correct. The core part of this process – the so-called view-warping process – is disparity compensation, where the position of every pixel in a view is shifted to that of another. Under a 1-D parallel camera setting for the simple case, the disparity shift applies only in the horizontal direction according to the given disparity values per pixel.

Likewise, it is important to determine an accurate disparity value, which can be obtained by the following equation:

$$d=\frac{f\cdot b}{Z} \quad (1)$$

where f is the focal length, b is the distance between the current and virtual views, and Z is the real depth of the object. Note that values in depth maps are scaled and quantized to be represented on a scale of 0 to L in the image. Therefore, the real depth value (Z) should be first obtained from the depth map sample value (D) to determine the disparity value, as in (2)

$$Z=\frac{L}{D(Z_{near}^{-1}-Z_{far}^{-1})+Z_{far}^{-1}} \quad (2)$$

where Z_{near} and Z_{far} indicate the actual distance from the nearest and farthest object in a scene, respectively, D represents the sample intensity of a depth map and L is the maximum value in a depth map (generally $L=255$). By (1) and (2), the relation between the disparity and the depth map sample values are given as

$$\begin{aligned} d&=\frac{fb}{L}\left[\left(Z_{near}^{-1}-Z_{far}^{-1}\right)D+Z_{far}^{-1}\right] \\ &=\alpha D+\frac{fb}{L}Z_{far}^{-1} \end{aligned} \quad (3)$$

In the equation, the parameter α is determined by the camera setting, and d is an affine function of D. This equation also aligns with common sense, i.e., an object closer to the screen and with a higher depth sample value records a larger disparity. It is known that distortion in the depth map causes that in the disparity value because it distorts the warping position. As illustrated in Fig. 2, an object with depth $Z=g(D)$ should have disparity d in the rendered view. However, a distortion in the depth map will yield a distorted real-depth value, $Z'=g(D')$, and will eventually cause the incorrect disparity (d'). Finally, it shifts the object by $\Delta d=d-d'$ in the rendered view.

2.2. Analysis of depth map distortion in 3DV system

The major difference between 2D and 3D video systems is the input and output data. Compared to a conventional