Signal Processing: *Image Communication* 41 (2016) 128–143

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Analysis of the depth-shift distortion as an estimator for view synthesis distortion

Pablo Carballeira*, Julián Cabrera, Fernando Jaureguizar, Narciso García

Grupo de Tratamiento de Imágenes, ETSI de Telecomunicación, Universidad Politécnica de Madrid, Ciudad Universitaria, 28040 Madrid, Spain

A R T I C L E I N F O

Article history:
Received 8 July 2015
Received in revised form
29 December 2015
Accepted 30 December 2015
Available online 8 January 2016

Keywords:
3D Video
Depth-image based rendering
Depth coding
Rate-distortion optimization

A B S T R A C T

Aiming for 3D Video encoders with reduced computational complexity, we analyze the performance of depth-shift distortion in depth-image based rendering algorithms, incurred when coding depth maps in 3D Video, as an estimator of the distortion of synthesized views. We propose several distortion models that capture (i) the geometric distortion caused by the depth coding error, (ii) the pixel-mapping precision in view synthesis and (iii) the method to aggregate depth-shift distortion caused by the coding error in a depth block. Our analysis starts with the evaluation of the correlation between the depth-shift distortion values obtained with these models, and the actual distortion on synthesized views, with the aim of identifying the most accurate one. The correlation results show that one of the models can be used as a reasonable estimator of the synthesis distortion in low complexity depth encoders. These results also show that the Sum of Absolute Error (SAE) captures better the distortion on a depth block than the Sum of Squared Error (SSE).

The correlation analysis is performed at three levels: frame, MB-row and MB. Results show that correlation values are consistently high at the frame level and for most MB-row positions, while lower values are achieved at the MB level and for specific rows at the MB-row level. Finally, to assess the results obtained by the correlation analysis, the different depth-shift distortion models are employed in two algorithms of the rate-distortion optimization (RDO) cycle of the depth encoder: (i) Quantization Parameter (QP) selection and (ii) mode decision. We evaluate the QP selection algorithm at three levels: frame, MB-row and MB, and the mode decision at the MB level. At the frame level, results show that the use of depth-shift distortion is equivalent to synthesis distortion, with the advantage of a lower computational complexity. At sub-frame levels, the results are consistent with the comparative correlation results, giving guidelines for the use of an efficient depth-shift distortion model on low complexity depth encoders.

1. Introduction

3D Video (3DV) and Free Viewpoint Video (FVV) are new visual media systems that expand the user's experience beyond what is offered by 2D video [1]. 3DV offers a 3D depth impression of the scene, while FVV allows an interactive selection of viewpoint and direction within a certain viewing range. To achieve those functionalities, a data format richer than a single 2D video signal is needed. The spectrum of data formats that can enable those functionalities goes from purely image-based data formats like multiview video (multiple views of the same scene) [2] to data formats based on computer graphics like 3D meshes and their corresponding textures [3]. A widely

* Corresponding author.
E-mail address: pcl@gti.ssr.upm.es (P. Carballeira).

http://dx.doi.org/10.1016/j.image.2015.12.007
0923-5965/

adopted approach is the one known as Multiview Video plus Depth (MVD), composed by *N* color sequences and *N* associated depth sequences, allowing the possibility of generating additional views on virtual camera positions [4]. However, the size of this MVD format grows linearly with the number of cameras that capture the scene and the available bandwidth of transmission systems is generally limited. Thus, an efficient compression scheme for the MVD format is needed. JCT-3V (collaborative project between ISO/IEC MPEG and ITU-T VCEG) is currently in the final steps of the standardization process of 3D-AVC [5], an extension of the H.264/MPEG-4 Advanced Video Coding (AVC) [6] standard that is specifically intended for this purpose.

When conventional block-based video encoders, such as AVC, or its multiview extension, Multiview Video Coding (MVC) [7] are used to encode depth, their target is to maximize the rate-distortion (RD) efficiency of coded depth sequences. Nevertheless, depth sequences are not seen by users, but used in virtual view synthesis. Lossy coding modifies the original values of depth map pixels, and such errors introduce a deviation in the pixel-mapping process of view synthesis algorithms. Therefore, the decisions on the encoder, (e.g. mode selection, motion estimation), should not be based on the visual quality of the decoded depth, but on the distortion on virtual views caused by the depth coding error. Some works have used view synthesis [8] or approximations of that view synthesis [9] to perform an efficient selection of encoding parameters in the depth encoder. Other works [10–12] described initial approaches to replace the conventional distortion function in depth coding for new distortion metrics that reflect this synthesis distortion. However, these approaches did not reflect completely the view warping/rendering process. In [13], the authors define a synthesis distortion metric and use it appropriately to optimize the coding of depth. This metric is used in the 3D-AVC encoder, enhancing the overall coding efficiency at the cost of a computational complexity overhead introduced by the new metric itself, and the fact that it requires joint processing of depth and texture in a single encoder.

With respect to the evaluation of the distortion in synthesized views, literature indicates that common video quality objective metrics are not sufficient to predict the quality of stereoscopic images and synthesized views [14]. However, most synthesis distortion metrics used to make encoding decisions in MVD encoders use common video quality metrics, i.e. PSNR. Although several efforts have been made to obtain reliable objective metrics to measure the quality of stereo pairs and synthesized views [15,16], a clear consensus over an adequate objective metric does not seem to have been reached yet. Regarding the use of common quality of synthesized views, some works show that metrics such as VQM or SSIM significantly outperform other objective metrics [17].

Here, we focus on low complexity depth encoders in real-time applications [18,19]. To avoid the computationally intensive synthesis at the depth encoder, it is desirable to predict view synthesis distortion directly from depth data characteristics. Therefore, we utilize a distortion model that uses the geometric distortion caused by the depth coding error as an estimation of the synthesis distortion. Lossy coding modifies the original values of depth map pixels, and such errors introduce a deviation in the pixel-mapping process of view synthesis algorithms. We refer to this deviation as *depth-shift distortion*. Also, the fact that pixels are mapped to integer positions (sub-pixel precision may be used but this is equivalent to frame upsampling with a later integer pixel-mapping) may affect the final depth-shift distortion value, thus it has to be taken into consideration too. Finally, as the macro-block (MB) is the coding unit in hybrid coding schemes such as AVC, different ways to aggregate the depth-shift distortion on a given depth block have also been analyzed. Therefore, we base our distortion model on: (i) the geometric distortion on view synthesis derived from the depth coding error, (ii) the pixel-mapping precision used in the view synthesis process and (iii) a method to aggregate the depth-shift distortion caused by the coding error in a depth block.

First, we analyze the effect of including the rounding on pixel-mapping and the aggregation methods on the computation of the depth-shift distortion through a statistical analysis of the correlation between different depth-shift distortion models and the actual distortion of synthesized views. For the aggregation of the depth-shift distortion on a given depth block we study two possibilities: Sum of Squared Error (SSE) and Sum of Absolute Error (SAE), to evaluate whether a quadratic or linear sum of pixel depth coding errors captures better the distortion on synthesized views. We perform the correlation analysis at three levels: frame, MB-row and MB, aiming to analyze the differences in the correlation between depth-shift and synthesis distortions at different sizes of depth coding units. This correlation analysis is supported by three different correlation metrics: Pearson, Spearman and Root Mean Square Error (RMSE). The results show high correlation of the depth-shift distortion and synthesis distortion at the frame and MB-row levels, with one of the depth-shift distortion models clearly outperforming the others, and indicating that it is beneficial to use SAE instead of SSE to aggregate depth coding error.

To assess the correlation results, we use the different depth-shift distortion models in the rate-distortion optimization (RDO) cycle, by means of two optimization algorithms: (i) a basic QP selection at different levels of the encoding process: frame, MB-row and MB; and (ii) mode decision at the MB level. Due to the lack of a unique objective quality metric that predicts the subjective quality of synthesized views, we evaluate the RD performance using several objective quality metrics. The RDO results are consistent with the correlation results, showing that the selected depth-shift distortion model is a reasonable estimator of the synthesis distortion for low complexity encoders at the MB-row level, while results at the MB level show a RD efficiency gap between the use of depth-shift and synthesis distortions.

This paper is organized as follows: in Section 2 we present our depth-shift distortion model, in Section 3 we describe the correlation analysis, and in Section 4 the RDO analysis. In Section 5 we present the conclusions.