BLOCK BASED RATE-DISTORTION ANALYSIS FOR QUALITY IMPROVEMENT OF SYNTHESIZED VIEWS

P. Carballeira,[†] G. Tech,[§] J. Cabrera,[†] K. Müller,[§] F. Jaureguizar,[†] T. Wiegand,[§] and N. García[†]

[†] Grupo de Tratamiento de Imágenes, ETSI de Telecomunicación,
Universidad Politécnica de Madrid, Ciudad Universitaria, 28040, Madrid, Spain.
[§] Image Processing Department, Fraunhofer Institute for Telecommunications
Heinrich Hertz Institute (HHI), Einsteinufer 37, 10587 Berlin, Germany.

ABSTRACT

We present a preliminary study on the Rate-Distortion (RD) gain that can be achieved applying RD optimization techniques in a multiview plus depth encoder. We consider the use of Multiview Video Coding (MVC) for both, color and depth sequences, and evaluate the improvement that can be obtained allowing a quantization parameter (QP) assignment on a macroblock basis compared to the use of a fixed QP for the whole sequence. The optimization criterion is the minimization of the distortion of the synthesized views generated at the receiver. Our motivation for this criterion is to capture the impact of depth coding according to its final purpose: the generation of virtual views. Since a unique objective quality metric for view synthesis artifacts evaluation has not been set yet, the performance of several algorithms for quality evaluation of the target synthesized view have been compared. Beyond obtaining a better RD performance, as could be expected, results also show that optimized synthesized views achieve absolute lower distortion values than the best result of the approach that uses a fixed QP for the whole sequence.

Index Terms — 3D video, multiview video coding, depth-image based rendering, rate-distortion optimization.

1. INTRODUCTION

The extension of 2D video to the third spatial dimension, providing the user with a depth impression of the observed scene, has acquired special relevance in the last years due to the development of 3D displays. Auto-stereoscopic displays allow the user to get a depth impression of the scene without the use of special glasses or head-set devices, and they also allow an interactive selection of real (captured by real cameras) or virtual viewpoints within the captured range.

Generally, 3D video systems use a data format that includes one or more video signals capturing the same scene (color signals from now on) and a geometrical representation of it. Different formats for scene geometry have been investigated [1][2], and one extensively used approach is the use of depth maps. These signals describe the distance of the scene objects to each of the cameras. One commonly used data format is the one known as Multiview

Video plus Depth (MVD), composed by N color sequences and N associated depth signals. The color and depth signals are jointly used at the receiver side to synthesize video signals corresponding to virtual viewpoints not captured by real cameras. The synthesized views can be displayed together with the original views on a variety of 3D displays.

Depth sequences can be seen as monochromatic video. Therefore, one efficient and backward compatible solution for this MVD format is to encode color and depth sequences independently into two separated MVC bitstreams [3]. However, frequency-domain characteristics of depth and color sequences are not equal and the quality of synthesized views is not addressed in the MVC encoder design. An optimal coding scheme shall maximize the relationship between the visual quality perceived by the user and the amount of transmitted data. In a 3D video scenario this implies that, in addition to the quality of the decoded original views, the quality of the synthesized views should also be considered in the encoder design and configuration.

The MPEG group is nowadays performing the preliminary exploration work for the development of the new 3D Video Coding (3DVC) standard [4]. This work includes the development of depth estimation and view synthesis algorithms, and a series of experiments analyzing the influence of the value of color/depth bitrate ratio in a MVC encoder, on the quality of synthesized views. However, this bitrate allocation experiments consider only a QP value assignment at the sequence level, as the only allowed configuration in the MVC reference software [5].

Block-based RD optimization techniques [6] have been applied to improve the performance of 2D video encoders. In a 3D video scenario, the traditional RD approach should be modified including not only the original views in the RD optimization target, but also the synthesized views. In this paper, we present an analysis of the RD gain that can be achieved through an approximation of a block-based RD optimization algorithm for MVD data sets in an MVC encoder. We have aimed our work only to the synthesized views, aligning it with the 3DVC group exploration work. Since no reference quality evaluation algorithm, which captures properly the visual impact of view synthesis artifacts, has been set yet, we have compared the performance of different quality evaluation algorithms in our RD optimization analysis. Results will show that beyond the expected bitrate saving, the quality of the synthesized views is improved compared to the best quality result obtained with a sequence-level selection of QP value.

The paper is organized as follows: in Section 2 we present the scheme of our 3D video RD optimization approach and in Sec-
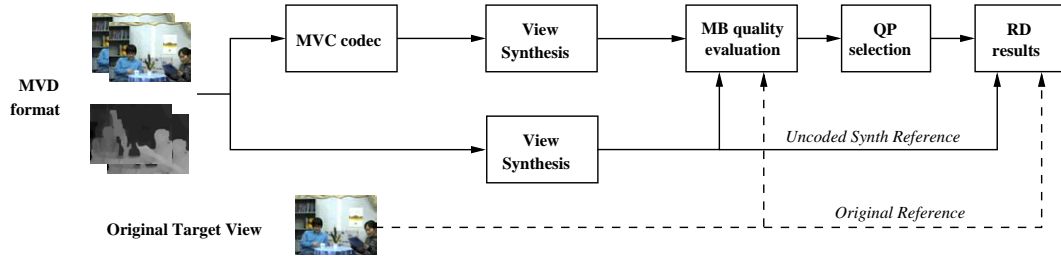
Figure 1. Scheme of our MB-based RD optimization approach for 3D video

tion 3 we present some implementation details and the RD performance results of the optimized coded MVD bitstreams.

2. APPROXIMATION OF 3D VIDEO RD OPTIMIZATION

The goal of our 3D video RD optimization approximation is to discuss the RD gain that can be achieved by means of RD optimization algorithms applied to color and depth data encoding. Our approach is based on an adaptive selection of the QP value of the MVC encoder on the macroblock (MB) level. This adaptive selection is made, independently, on both color and depth MVC encoders, based on the quality of a target MB of the synthesized view, and not the quality of the decoded color or depth data. The effects of color and depth quantization on synthesis quality differ from coding effects on 2D video, in which lower QP values result in a higher decoded quality. The results of our RD optimization approach will show that, beyond bitrate saving, the quality of the resulting optimized synthesized view is increased with respect to the synthesized view with highest quality among all possible solutions with QP assignmet at sequence level.

The scheme of our approach is shown on Figure 1. The original MVD data is encoded and decoded using the MVC scheme and a fixed QP for the whole sequence. This process is repeated for different QP values resulting in different values of bitrate. The decoded color and depth data sets are then used to generate different samples of the synthesized view for all the different color and depth bitrate possible combinations. Then, the objective quality of each MB of each sample is evaluated relative to the uncoded synthesized view or to the original target view. Finally, for each MB position, the corresponding MB with higher quality among all samples is selected and an optimized synthesized view is composed. The quality and bitrate of the resulting view is compared with the fixed-QP synthesis results.

Our approach has to be considered as an approximation of a real RD optimization algorithm since the process is done offline and the following simplifications were considered:

- A simple RD optimization criterion was used: the combination of color and depth QPs, that lead to the best MB synthesis quality, is always selected

- Only the quality of one single synthesized view is considered in the RD target

- The bitrate of the optimized stream is computed using an approximation described in 2.2

Despite these simplifications, the results might be considered a lower bound for the possible RD gain, since higher RD improvement ratios could be achieved by means of more complex RD optimization processes.

2.1. Objective video quality evaluation

2D picture quality evaluation methods are not the most adequate to measure 3D image quality since typical stereoscopic artifacts (e.g. crosstalk) are not incorporated. Up to now, the definition of an objective quality measure, reliably predicting the perceived quality of 3D sequences, has not been solved, and only a few objective 3D quality evaluation methods have been put forward [7].

In addition to stereoscopic artifacts, view synthesis artifacts are also visible on synthesized views even when only one synthesized 2D sequence is displayed in a regular 2D display. A reliable measure that incorporates properly the visual effect of this type of artifacts has not been set yet. In our work, we have tested the following algorithms to evaluate the quality of the MBs of the synthesized views:

- Peak Signal-to-Noise Ratio (PSNR)

- Video Structural SIMilarity (VSSIM) [8]

- Spatial Peak Signal to Perceptual-Noise Ratio (SPSPNR)[9]

Compression of color and depth data prior to view synthesis results in additional visual artifacts. Furthermore, coding and view synthesis artifacts are related, e.g. coding artifacts in the depth data results in view synthesis artifacts. The 3DVC group in MPEG has made an effort in depth estimation and view synthesis optimization, to achieve high quality uncoded synthesized views so that view synthesis and coding artifacts can be decoupled. If the uncoded synthesized view is used as reference, only the coding artifacts are evaluated while if the original view is used as reference, also view synthesis artifacts are included in the distortion measure. In our work we have used both references when available.

2.2. Macroblock-wise QP selection

For each color and depth QP combination, a sample of the synthesized view is generated. Given that set of samples and the quality value for each MB of each sample, the following process was performed:

- For each MB position, the MB of the synthesized sample with the best quality is selected.

- The number of bits of each MB is evaluated by averaging the number of bits of the MBs involved in the 3D warping process.

- The optimized synthesized sequence is built as the union of the selected MBs

- The RD performance of the optimized synthesized sequence is compared against the RD performance of the fixed-QP synthesis results
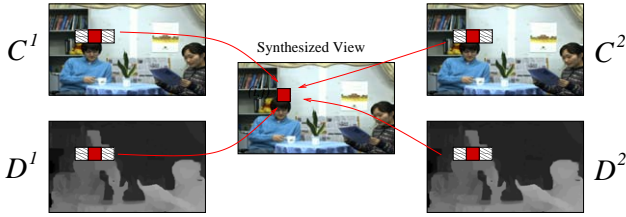
Figure 2. Computation of the number of bits of a MB

In this analysis, the number of bits dedicated to the MB of the synthesized view is computed under the assumption of a parallel configuration of cameras, and therefore, a unidimensional view synthesis algorithm, for which only horizontal displacement of pixels is permitted. The number of bits of MB with position (i, j), is computed using the following expression:

$$Bits(i,j) = \frac{1}{3} \sum_{k=1}^{N} \sum_{l=i-1}^{i+1} (C_{l,j}^k + D_{l,j}^k), \qquad (1)$$

where N is the number of views considered in the view synthesis algorithm, and $C_{l,j}^k$ and $D_{l,j}^k$ are the number of bits of the MB with position (l, j) of view k for color and depth respectively in the MVC bitstreams. That is, the number of bits of the MVD coded representation corresponding to a given MB is obtained averaging the number of bits of three adjacent MBs on the color and depth sequences on the horizontal dimension. Figure 2 illustrates an example of the later bit computation.

3. EXPERIMENTAL RESULTS

The RD optimization process described in Section 2, has been applied to *Newspaper* and *Beergarden* sequences, two data sets from the 3DVC reference set. Both sequence have two views. While for *Beergarden* the depth content is synthetic and the original target view is not available, for *Newspaper* the depth content was obtained by depth estimation and the original target view is available.

3.1. Multiview coding of MVD data

The original color sequences, as well as the corresponding depth sequences, were encoded jointly using the MVC reference software (JMVC 5.0.5). Color and depth sequences were encoded in two separated MVC bitstream using a hierarchical B-frame GOP such as the one described in [10]. The parameters of the MVC encoder configuration are shown in Table 1.

3.2. View synthesis

The decoded color and depth signals were used for the view synthesis of a virtual view located in the middle point between both original cameras. The possible combinations of color and depth QP choices add up to 40 (5 for color and 8 for depth, see Table 1). We have used the MPEG view synthesis reference software VSRS 3.5 [11]. This software implements a 3D warping algorithm [12] restricted to the horizontal dimension (the cameras are set in a parallel configuration) and several artifact enhancement techniques, such as bilateral filtering, adaptive image blending and boundary matting.

Table 1. Values of the MVC configuration parameters

GOP size	15 frames
Search range	96 pixels
Entropy encoding	CABAC
Color QPs	22, 26, 30, 34, 38
Depth QPs	22, 26, 30, 34, 38, 42, 46, 50

3.3. RD results

Figure 3 shows the results of the RD optimization framework for both test sequences and the three different quality evaluation algorithms relative to the available reference views. In each graph, the RD performance of the optimized bitstream (red point) and the fixed-QP combinations (blue points) can be compared. Results show that the RD optimization approach is valid for all the tested cases, since the RD performance of the optimized bitstream is better than each of the fixed-QP combinations. Furthermore, not only bitrate is saved, but the quality of the optimized synthesized views is higher than the highest quality obtained with fixed-QP combinations. In the absence of a unique reliable quality evaluation algorithm for synthesized views, the consistent results for all the tested quality metrics indicate the goodness of our RD optimization approach. The RD results using the uncoded synthesized view as reference are closer to a real implementation, since the original target view may not be available in a real MVD encoder. Among the three quality evaluation algorithms, the best RD improvement is obtained for the SPSPNR measure. However, the adequateness of the objective evaluation algorithms to capture the quality perception of the user should be evaluated by subjective assessment of the results.

4. CONCLUSIONS

This paper presented a preliminary study on the RD gain that can be achieved by means of optimization techniques in a MVD data encoder. Through our RD optimization approximation we have shown that bitrate saving and even objective quality gain for synthesized views can be achieved by means of a QP selection at the MB level. The consistent results for different quality evaluation algorithms and test data confirm the advantages of our approach. The results encourage the use of RD optimization algorithms in real MVD data encoders. Despite our focus on the MVC scheme, RD optimization algorithms do not necessarily need to be restricted to a specific coding scheme and could also be applied to future block-based coding schemes for MVD data. Possible future implementations could include more complex RD optimization processes to increase the RD improvement.

5. REFERENCES

[1] W. Matusik, C. Buehler, and L. McMillan, "Polyhedral visual hulls for real-time rendering," in *Proceedings of 12th Eurographics Workshop on Rendering, Eurographics Association*, December 2001, pp. 115–125.

[2] S. Würmlin, E. Lamboray, and M. Gross, "3D video fragments: dynamic point samples for real-time free-viewpoint video," *Computers and Graphics, Special Issue on Coding, Compression and Streaming Techniques for 3D and Multimedia Data*, vol. 28, no. 1, pp. 3–14, 2004.
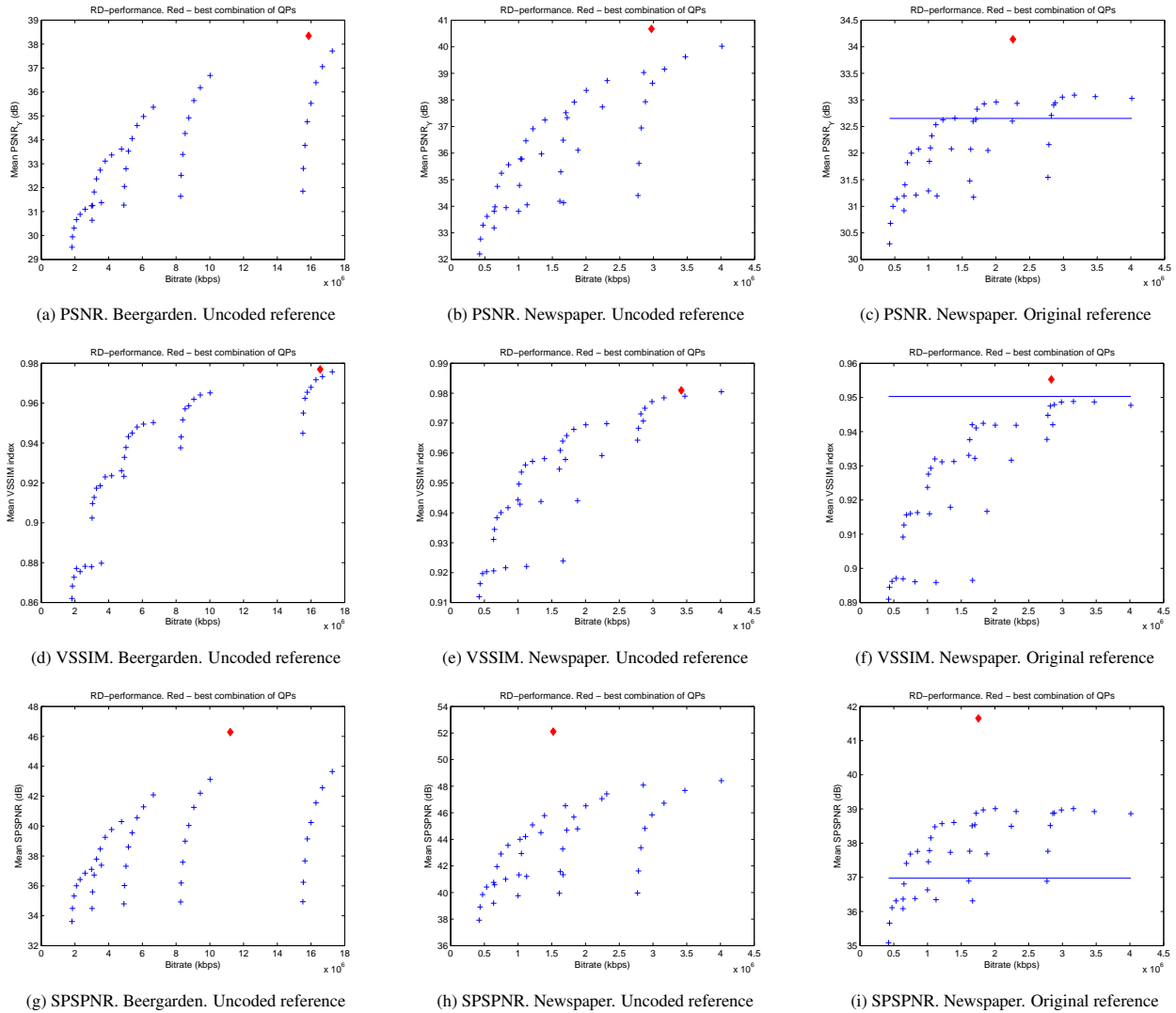
Figure 3. RD performance of the optimized coded MVD bitstreams (red) and fixed QP combinations (blue) for different quality evaluation algorithms. The objective quality of the synthesized view is plotted against the bitrate of the coded MVD bitstream. In the graphs for the original reference the horizontal line shows the quality of the uncoded synthesized view compared to the original.

[3] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proceedings of IEEE International Conference on Image Processing*, October 2007, pp. 201–204.

[4] ISO/IEC JTC1/SC29/WG11, "Vision on 3D video," *output doc. N10357*, Laussane, Switzerland, February 2009.

[5] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-5:2001/FDAM 15 MVC Reference Software," *output doc. N10897*, Xian, China, October 2009.

[6] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74–90, November 1998.

[7] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntiens, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 381–391, March 2004.

[8] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication, special issue on objective video quality metrics*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[9] Y. Zhao and L. Yu, "Perceptual measurement for evaluating quality of view synthesis," *ISO/IEC JTC1/SC29/WG11, contrib. M16407*, Maui, USA, April 2009.

[10] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y. Wang, "Joint Multiview Video Model (JMVM) 8.0," *output doc. N9762*, Archamps, France, April 2008.

[11] Y. Zhao, L. Yu, and D. Tian, "Improved 1D mode in VSRS 3.1," *contrib. M16582*, London, UK, July 2009.

[12] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems*, January 2004, vol. 5291, pp. 93–104.