



HAL
open science

A study of depth/texture bit-rate allocation in multi-view video plus depth compression

Emilie Bosc, Fabien Racapé, Vincent Jantet, Paul Riou, Muriel Pressigout, Luce Morin

► **To cite this version:**

Emilie Bosc, Fabien Racapé, Vincent Jantet, Paul Riou, Muriel Pressigout, et al.. A study of depth/texture bit-rate allocation in multi-view video plus depth compression. *Annals of Telecommunications - annales des télécommunications*, 2013, 68 (11-12), pp.615-625. <10.1007/s12243-013-0363-x>. <hal-00907792>

HAL Id: hal-00907792

<https://hal.science/hal-00907792v1>

Submitted on 21 Nov 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A study of depth/texture bit-rate allocation in Multi-View Video plus Depth compression

Emilie Bosc · Fabien Racapé · Vincent Jantet · Paul Riou · Muriel Pressigout · Luce Morin

Received: date / Accepted: date

Abstract Multi-view Video plus Depth (MVD) data offer a reliable representation of three dimensional (3D) scenes for 3D Video applications. This is a huge amount of data whose compression is an important challenge for researchers at the current time. Consisting of texture and depth video sequences, the question of the relationship between these two types of data regarding bit-rate allocation often raises.

This paper questions the required ratio between texture and depth when encoding MVD data. In particular, the paper investigates the elements impacting on the best bit-rate ratio between depth and color: total bit-rate budget, input data features, encoding strategy, assessed view.

Keywords Multi-view video coding · bit-rate · HEVC · H.264 · PSNR · view synthesis

1 Introduction

3D video applications rely on exhaustive 3D representations of real or synthetic scenes. Among the numerous possible representations [1] are the Multi-View Video plus Depth (MVD) sequences. They consist of multiple texture video sequences coupled with depth video sequences, acquired at different viewpoints of the represented scene. The question of MVD compression is delicate because artifacts in 3D video sequences raise new and unsolved issues related to human vision mechanisms [2].

Emilie Bosc, Fabien Racapé, Muriel Pressigout, Luce Morin
Université Européenne de Bretagne, INSA de Rennes, IETR, UMR 6164, 35708, Rennes, France
E-mail: firstname.name@insa-rennes.fr

Vincent Jantet

Both 3D video applications, namely Free Viewpoint Video (FVV) and 3D Television (3DTV), require the synthesis of novel viewpoints in order to stimulate human vision system and to simulate parallax motion. This is achieved thanks to Depth Image Based Rendering (DIBR) [3] algorithms that use both texture and depth data. Since the quality of the synthesized views depends on the contribution of both texture and depth data, the issue of bit rate allocation between these two types of data must be addressed when designing an MVD coding framework.

Many approaches proposed a solution to the depth/texture bit-rate allocation issue but the optimal ratio criterion remains open. At the beginnings, the objective quality of the depth map was thought to be comparable to that of the color image. Then by the same token, the color signal having three components and the depth signal having only one component, it seemed appropriate to allocate 25% of the total bit-rate to depth data. Besides, in [3], Fehn *et al.* show that being a gray-scale signal, the depth video can be compressed more efficiently than the texture video and recommend using less than 20% of the texture bit-rate for video-plus-depth data format. This recommendation is based on the fact that “*the per-pixel depth information doesn’t contain a lot of high frequency components*” [3]. In this approach proposed by the AT-TEST project [3, 1], a base layer encodes a conventional 2-D colour video using MPEG-2 and an advanced layer encodes the depth information.

Yet, depth maps define the very structure of the scene. Artifacts occurring in critical areas of the depth map may lead to fatal synthesis errors when generating virtual views. Critical areas are known to be the sharp depth planes edges (i.e. high frequencies) whose encoding is costly and particularly prone to errors. To address this problem other studies [4] proposed to consider the distortion of the depth information. In [4], Daribo *et al.* proposed a coding method for the video-plus-depth data through a joint estimation of the MV field for the texture motion information and the depth map sequence. This coding method also includes a bit-rate allocation strategy between the texture and depth map stream based on a rate-distortion criterion. Both the distortion (Mean Squared Error, MSE) of texture and depth data are taken into account in the proposed model of [4]. However, a previous study [5] also shown that the distortion of depth maps in terms of MSE may not be appropriate to predict the quality of the resulting synthesized views. In addition, since the depth map will not actually be observed by the users of an 3D Video application, it seems more pertinent to consider the quality of the synthesized views, that are more likely to be observed.

Thus, most of the up-to-date bit-rate allocation strategies now rely on the quality of the synthesized views. This is motivated by the realization of the impact of depth map compression on visual quality of the synthesized views ([6], [7]). Different bit-rate allocation strategies have been proposed in the literature. Some of them opt for the fixed-ratio bit-allocation scheme, like Martinian *et al.* [8] who consider using only 5-10% of the total bit-rate for encoding depth

data with H.264/AVC, based on the observation that it “*produced acceptable view synthesis performance*”.

Other strategies [9,10,11] rely on a global optimization equation solver based on the objective image quality of the rendered intermediate view. In their experiments [9], Ekmekcioglu *et al.* consider a given total budget where 15% is assigned to depth data. The authors noticed that their proposed bit-allocation strategy based on the synthesized view quality, outperforms the conventional constant-ratio allocation strategy.

Similarly, in [10], the authors proposed an efficient joint texture/depth rate allocation method based on a view synthesis distortion model, for the compression of MVD data. According to the bandwidth constraints, the method delivers the best quantization parameters combination for depth/texture sets in order to maximize the quality of the rendered synthesized view in terms of MSE (Mean Squared Error) (with respect to the original captured view). The proposed model finds the optimal ratio between depth bit-rate and texture bit-rate. However, the optimization depends on the target virtual viewpoint.

In [11], the authors proposed a bit-rate allocation strategy between texture and depth maps of multi-view images based on a model that minimizes the distortion of the synthesized views. A cubic distortion model based on basic DIBR properties drives the optimal selection of coded views and quantization levels for corresponding texture and depth maps. Cheung *et al.* ([11]) observe improvements by up to 1.5 dB with this proposed model, compared to constant-ratio allocation strategy.

The 3D-HTM coding software under standardization of MPEG includes a depth coding tool called View Synthesis Optimization (VSO) [12] that is meant to improve depth bit-rate allocation based on the quality of a synthesized view. This tool integrates a distortion term calculated based on the quality of texture synthesized using decoded depth.

This previous work on bit-rate allocation between texture and depth data highlights the fact the challenges raised by this issue. These challenges are not that trivial since they seem to involve many questions:

- the importance of the preservation of depth sharp edges (discontinuities)
- the influence of the target intermediate synthesized viewpoint used for the optimization process
- other questions such as the bit-rate allocation within one frame, or between frames

Based on our previous studies [13,14], we rely on the assumption that there exists an optimal ratio between texture and depth data. The experiments presented in this paper investigate the factors impacting on the optimal ratio between texture and depth data in MVD compression. In particular, we question the role of 1) the total bit-rate budget, 2) the coding method, 3) the input data features and 4) the location of the assessed view with respect to the decoded reference views. This paper aims at determining the impact of these four aspects in the bit-rate allocation between texture and depth data, in the con-

text of MVD data compression in FVV applications by using objective metrics.

The study includes the compression of MVD data sequences at different bit-rates in order to determine the best bit-rate distribution between depth and texture, when based on Peak Signal to Noise Ratio (PSNR) measures of the synthesized view. In most of the studies, the use of this objective metric is justified by its simplicity and mathematical easiness to deal with such purposes. Thus, we also investigate the bit-rate allocation issue when based on such a metric.

The paper is organized as follows: Section 2 presents the experimental protocol basis. The next sections present the results of our investigations depending on the studied element assumed to impact on the bit-rate between texture and depth data: Section 3 addresses the analysis of the coding constraints; Section 4 addresses the influence of the MVD sequences features and of the assessed viewpoint. Finally Section 5 concludes the paper.

2 Global experimental conditions

The following experiments aim at investigating the factors impacting the best ratio between texture and depth data, when encoding MVD data, in a FVV application context. For this reason, the experiments are both based on the same global experimental conditions, as depicted in Fig. 1. Texture and depth data from MVD sequences are processed by an encoding method. The reconstructed views are used for generating an intermediate viewpoint thanks to the View Synthesis Reference Software (VSRS) version 3.5, provided by MPEG. The synthesized viewpoint is then evaluated relying on the PSNR with respect to the original acquired view. Depending on the studied criterion (coding method, data features, etc...), some parameters slightly differ from the global experimental conditions. The details are described in the concerned sections. In the following, we present the results of our investigations regarding the factors influencing the best ratio between texture and depth data.

3 Analysis of coding constraints

In this section, the effects of the encoding method constraints are studied. For this purpose, following the basis scheme described in Section 2, H.264/MVC reference software, JMVM 8.0 (Joint Multiview Video Model) [15] is used to encode three views, as a realistic simulation of a 3D-TV use. To vary the bit-rate ratio and the total bit-rate, the quantization parameter QP varies from 20 to 44 for both depth and texture coding. The central view is used to predict the two other views in the case of H.264/MVC. Then, from the decompressed views, we computed an intermediate view between the central view and the right one, by using the reference software: VSRS, version 3.5, provided by

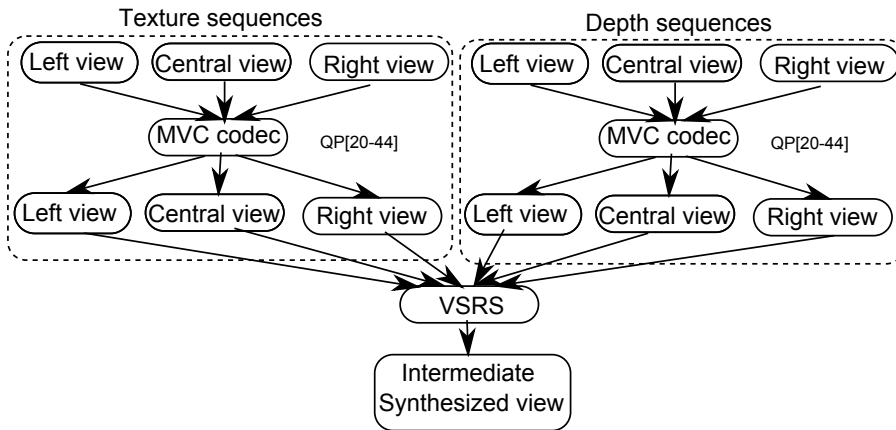


Fig. 1 Global experimental conditions.

MPEG.

In this case study, “MVC codec” in Figure 1 thus refers to H.264/MVC.

We used two different types of sequences to answer our question: *Ballet* from Microsoft Research, and *Book Arrival* from Fraunhofer HHI (1024×768). This last sequence is a 3DV test material in MPEG.

The considered views are 2, 4 and 6 for *Ballet*, and 6, 8 and 10 for *Book Arrival*. Viewpoint 3 is generated from decoded viewpoints 2 and 4 in the case of *Ballet*. Viewpoint 9 is generated from decoded viewpoints 8 and 10 in the case of *Ballet*. For each couple $(QP_{texture}, QP_{depth})$, the average PSNR score of the synthesized sequence is evaluated, compared to the original acquired view. The obtained results are discussed in the following in an organized manner, according to the investigated factors.

| | $QP_{texture}$ | QP_{depth} |
|-------|--|--|
| H.264 | 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44 | 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44 |
| HEVC | 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44 | 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44 |

3.1 Visual analysis

Here, we address the visual image quality of the rendered views synthesized from the decompressed data. Fig. 2 gives selected snapshots of the resulting synthesized views when using H.264-decompressed data with various texture/depth ratios.

Figures 2(b), 2(f) and 2(j) show that allocating less than 10% of the total bit-rate to depth data induces important damages along the edges. The location of the depth map discontinuities is deeply compromised which leads to errors in the synthesized view. PSNR scores fall down because of the numerous errors along the contours of objects.

Figures 2(d), 2(h) and 2(l) suggest that affecting more than 80% of the total bit-rate to depth data preserves the edges of some objects but texture information is lost because of the coarse quantization.

Affecting between 40% and 60% of the total bit-rate to depth data seems to be a good trade-off for the tested sequences, as it can be observed in Figure 2(c), 2(g) and 2(k). PSNR scores and visual quality are both improved compared with the two other presented cases. The depth maps are accurate enough to ensure correct projections and decompressed texture images are good enough to avoid drastic artifacts.

The obtained results showed that the best quality of reconstruction by using VSRS may require to affect between 40% and 60% of the total bit-rate to depth data, depending on the available MVD data.

3.2 Dependence on the total bit-rate

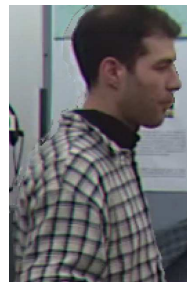
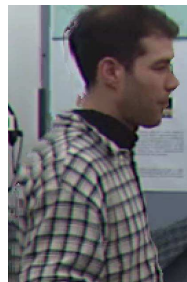
Fig. 4 presents the interpolated rate-distortion curves of synthesized views for *Ballet* sequence and *Book Arrival* sequences using H264/MVC. The goal of these figures is to visualize the allocated depth ratio obtaining the highest image quality in terms of PSNR. For this reason, these figures plot the PSNR values on the ordinate and the ratio allocated to depth, in percentage, with respect to the total bit-rate (texture+depth), on the abscissa. The obtained synthesized views PSNR scores have been grouped into bit-rate intervals. One color corresponds to one total bit-rate interval (texture+depth), in the figures. In the figures, points correspond to the actual measured scores. Dots with same color belong to same bit-rate range interval. The different curves correspond to interpolation of the measured points for each range of bit-rate.

Our results are consistent with [16]: although the authors do not state clearly the appropriate ratio for those two sequences, their rate/distortion curves show that, for example, the bit-rate pair (962kbps for texture, 647kbps for depth), i.e. a percentage of 40% for depth, gives better synthesis quality (in terms of PSNR) for *Book Arrival*. The synthesis conditions are similar to our experiments. On the other hand, in [3], the synthesis conditions involve one single video-plus-depth data: in this case, a very little continuum is supported around the available original view. Since synthesis distortion increases with the distance of the virtual viewpoint, this explains the significant difference with our results.

We observe that for a given sequence, no matter the bit-rate, the ratio that provides the best quality seems relatively constant: around 60% for *Ballet*, and around 40% for *Book Arrival*. However, a close observation to the curves shows that this bit-rate ratio lies in a narrow interval of values that depends on the total bit-rate. The value of the inflection points slightly decreases when the total bit-rate decreases.



(a) Synthesis from original data, PSNR = 37.54 dB
 (b) PSNR = 30.0dB; Depth = 3% of bit-rate;(0.19bpp in total)
 (c) PSNR = 33.8dB; Depth = 60% of bit-rate;(0.12bpp in total)
 (d) PSNR = 30.8dB; Depth = 95% of bit-rate;(0.13bpp in total)



(e) Synthesis from original data, PSNR = 40.1dB
 (f) PSNR = 36.96dB; Depth = 6% of bit-rate;(0.21bpp in total)
 (g) PSNR = 39.38dB; Depth = 38% of bit-rate;(0.21bpp in total)
 (h) PSNR = 34.17dB; Depth = 88% of bit-rate;(0.09bpp in total)



(i) Synthesis from original data, PSNR = 40.1dB
 (j) PSNR = 36.96dB; Depth = 6% of bit-rate;(0.21bpp in total)
 (k) PSNR = 39.38dB; Depth = 38% of bit-rate;(0.21bpp in total)
 (l) PSNR = 34.17dB; Depth = 88% of bit-rate;(0.09bpp in total)

Fig. 2 Synthesized images from H.264 encoded MVD data, with different bit-rate ratios between texture and depth. a), b), c) and d) refer to the sequence *Ballet*. e), f), g), h), i), j) k) and l) refer to the sequence *Bokk Arrival*.



(a) Synthesis from original data, PSNR = 37.54 dB
 (b) PSNR = 36.22dB; Depth = 3% of bit-rate;(0.08bpp in total)
 (c) PSNR = 37.36dB; Depth = 60% of bit-rate;(0.06bpp in total)
 (d) PSNR = 36.3dB; Depth = 89% of bit-rate;(0.08bpp in total)



(e) Synthesis from original data, PSNR = 40.1dB
 (f) PSNR = 38.99dB; Depth = 6% of bit-rate;(0.034bpp in total)
 (g) PSNR = 38.02dB; Depth = 38% of bit-rate;(0.008bpp in total)
 (h) PSNR = 35.63dB; Depth = 88% of bit-rate;(0.01bpp in total)



(i) Synthesis from original data, PSNR = 40.1dB
 (j) PSNR = 38.99dB; Depth = 6% of bit-rate;(0.034bpp in total)
 (k) PSNR = 38.02dB; Depth = 38% of bit-rate;(0.008bpp in total)
 (l) PSNR = 35.63dB; Depth = 88% of bit-rate;(0.01bpp in total)

Fig. 3 Synthesized images from HEVC encoded MVD data, with different bit-rate ratios between texture and depth. a), b), c) and d) refer to the sequence *Ballet*. e), f), g), h), i), j), k) and l) refer to the sequence *Book Arrival*.

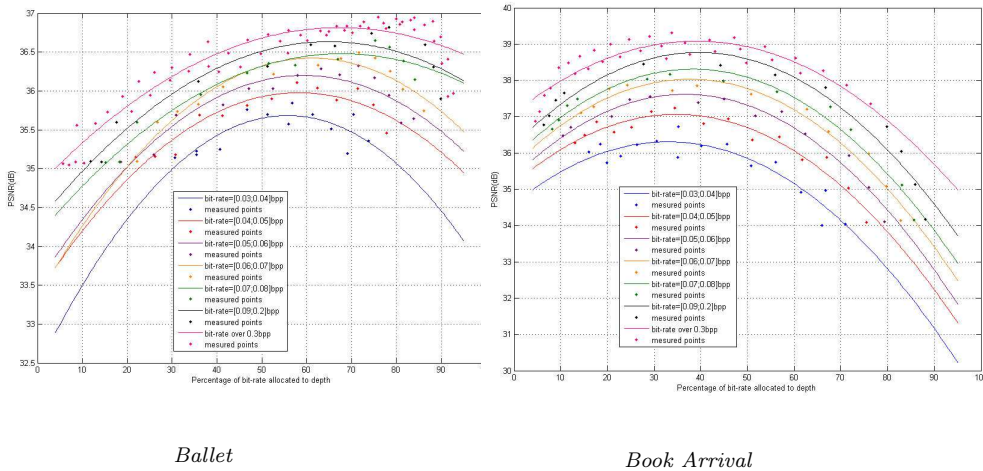


Fig. 4 Interpolated rate-distortion curves of synthesized views (from H.264 encoded MVD data). PSNR (dB) of synthesized views as a function of rate allocated to depth in percentage of total rate.

3.3 Dependence on the coding method

The previous subsections observations (3.2 and 3.1) are related to H.264/MVC encoding. Using a different encoding framework may lead to a different ratio for depth. For this reason, our experiment includes another compression method (HEVC), in the same test conditions. This subsection presents the results.

As in the previous experiment using H.264/MVC, we observe the visual image quality of the rendered views, from the decompressed data, as depicted in Fig. 3. The analysis of Fig. 2 and 3 leads to interesting observations regarding the influence of the encoding method on the bit-rate ratio between texture and depth data.

For similar ratios, the visual quality of the synthesized views is different depending on the used encoding method: in the case of H.264/MVC encoding, Figures 2(c), 2(g) and 2(k) show a correct visual quality when allocating around 60% and 40% (for *Ballet* and *Book Arrival* respectively to depth data. However, in the case of HEVC, this compromise is not appropriate since Figures 3(c), 3(g) and 3(k) show inaccurately rendered edges. The observation of Figure 3 suggests that the optimal depth ratio is in a range from 3% to 60% for *Ballet* and around 6% for *Book Arrival*, with HEVC. These comments suggest that the encoding method impacts on the optimal ratio between texture and depth data considering visual image quality observations. An important difference between H.264 and HEVC lies in the replacement of H.264 macroblocks by flexible coding scheme based on coding units of variable size (from 64×64 to 8×8 blocks) that can be hierarchically subdivided (through a quad-tree subdivision). In addition, as previously studied in [17] the gain also comes from changes as the improvement of intra prediction modes and directions and the use of asymmetric partitions for inter prediction. The results thus

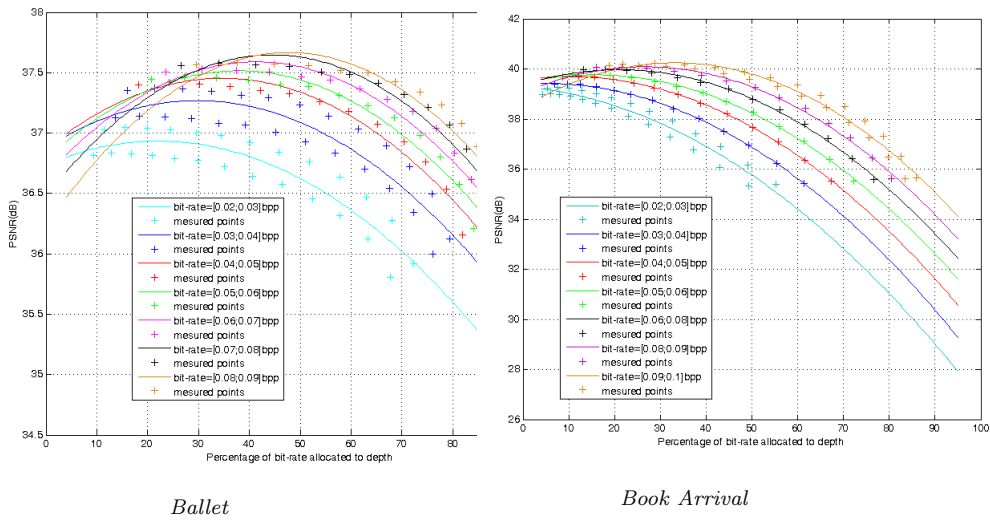


Fig. 5 Interpolated rate-distortion curves of synthesized views (from HEVC encoded MVD data). PSNR (dB) of synthesized views as a function of rate allocated to depth in percentage of total rate.

suggest that HEVC representation is more adapted for depth map compression than H.264, considering the visual observations of the synthesized views respectively: despite a lower rate allocated to depth data, essential depth information is preserved to enhance the quality of the synthesized view in terms of PSNR, in these experiments. As expected, we also note that the visual artefacts in the presented synthesized views differ depending on the coding strategy.

Fig. 5 depicts the interpolated rate-distortion curves of synthesized views for *Ballet* sequence and *Book Arrival* sequences using HEVC. In average, with HEVC, the percentage of bit-rate allocated to depth data leading to the maximum PSNR is 27.5% for *Ballet* and 12.2% for *Book Arrival*. In the case of HEVC, the obtained curves show that the optimal ratio depends on the total bit-rate budget, while remaining in a narrow range: the inflection point values of the curves decreases when the total bit-rate budget decreases. The obtained ranges are [21%-34%] for *Ballet* and [9%-15%] for *Book Arrival*. This suggests that the lower the bit-rate budget, the less depth data is required, which implies a necessary return to 2D at very low bit rates.

4 Analysis of data characteristics

In this section, the effects of the influence of the MVD sequences features and of the assessed synthesized view are studied. In particular, we assume that the video content, the complexity of depth and the camera settings are related to the bit allocation between depth and texture. The following experiments are in line with this concern. The basis experimental scheme described in 2

now includes H.264/MVC codec only with 11 sequences. For each sequence, we encoded different frames. Then we generated the virtual viewpoints with various baseline distances between the reference viewpoints. Table 1 gives the summary of the used material. Table 2 summarizes the tested sequences features. This table is partly based on the analysis provided in [18]. The encodings follow the basis scheme protocol described in 2: left and right views (textures and depth maps) are encoded through MVC reference software (JMVM 8.0). Based on the same protocol, the optimal ratio between depth and texture are calculated, with PSNR of an intermediate synthesized viewpoint as an indicator of distortion.

Table 3 summarizes these results. The obtained bit-rate ratios vary from 16% to 52% depending on the test sequence. The analysis of this table, compared to Table 2 does not allow a clear conclusion of the influence of the sequences characteristics on the bit-rate ratio. For this reason, we propose a further study including new axes to be investigated. They are the features which differ from one content to another: depth map accuracy, depth structure complexity, baseline distance between the reference cameras, features of discovered areas. These aspects are addressed by three analyses in the following: depth maps entropy, baseline distance between the reference cameras, high contrast between background and foreground around the discovered areas.

4.1 Depth maps entropy and texture images entropy

We assume that the ratio which rules the optimal synthesized views in terms of PSNR, is related to the amount of information contained in the original data. In other words, the entropy of depth against the entropy of texture is expected to influence the optimal allocation between depth and texture. Let e_d be the average entropy of the encoded depth maps, for a given content. Let e_t be the average entropy of the encoded texture frames, for the same content. For each tested content, we computed the following ratio:

$$R_e = \frac{e_d}{e_d + e_t} \quad (1)$$

Fig. 6 plots the computed mean R_e of per sequence against the “optimal” percentage of bit-rate allocated to depth data according to our previous experimental protocol. There is relationship between R_e and the “optimal” percentage of bit-rate allocated to depth data. The correlation coefficient between R_e and the “optimal” percentage of bit-rate allocated to depth data was calculated through a Matlab-made Pearson coefficient implementation and it reached 76.95%. These results are understandable because a high entropy value for the depth implies a highly detailed depth structure. If the level of details

| Sequence Name | Frame no. | Left and Right views | Central view |
|---------------|-----------|----------------------|--------------|
| Ballet | 1 | 0-2 | 1 |
| | 100 | 0-2 | 1 |
| Balloons | 1 | 1-3 | 2 |
| | | 1-5 | 3 |
| | | 3-5 | 4 |
| | 10 | 1-5 | 3 |
| | 50 | 1-5 | 3 |
| | 300 | 1-3 | 2 |
| | | 1-5 | 3 |
| 3-5 | | 4 | |
| Book Arrival | 1 | 8-10 | 9 |
| | 99 | 8-10 | 9 |
| Breakdancers | 1 | 0-2 | 1 |
| | | 0-4 | 1 |
| | | | 2 |
| | | | 3 |
| | | 0-6 | 1 |
| | | | 3 |
| | | 0-7 | 1 |
| | 4 | | |
| | 100 | 1-4 | 3 |
| | | 2-6 | 4 |
| | | 4-6 | 5 |
| 4-7 | | 6 | |
| Cafe | 1 | 0-2 | 1 |
| | 300 | 2-4 | 3 |
| Champagne | 1 | 2-4 | 3 |
| | 300 | 2-4 | 3 |
| Champagne | 1 | 37-41 | 39 |
| | 300 | 37-41 | 39 |
| Kendo | 1 | 1-3 | 2 |
| | | 1-5 | 2 |
| | | | 3 |
| | 300 | 3-5 | 4 |
| | | 3-5 | 4 |
| Pantomime | 1 | 37-41 | 39 |
| | 500 | 37-41 | 39 |
| Mobile | 1 | 3-5 | 4 |
| | | 3-7 | 5 |
| | | 3-7 | 6 |
| | 100 | 3-5 | 4 |
| | | 3-7 | 5 |
| | | 3-7 | 6 |
| | 200 | 3-5 | 4 |
| 3-7 | | 5 | |
| Lovebird | 1 | 3-7 | 6 |
| | | 4-8 | 6 |
| Newspaper | 1 | 2-4 | 3 |
| | | 2-6 | 3 |
| | | | 4 |
| | | | 5 |
| | 2 | 4-6 | 5 |
| | 10 | 2-6 | 4 |
| | 50 | | |
| 300 | | | |

Table 1 Test material.

| Sequence Name | Characteristics | Depth structure complexity | Camera spacing |
|---------------|--|----------------------------|---|
| Ballet | natural scene, high detail | high | varying, toed-in configuration |
| Balloons | natural scene, moving cameras, high detail | high | stereo distance, parallel configuration |
| Book Arrival | natural scene, high detail | high | stereo distance, parallel configuration |
| Breakdancers | natural scene, high detail | high | varying, toed-in configuration |
| Cafe | natural scene, medium detail | medium | stereo distance, parallel configuration |
| Champagne | natural scene, high detail | medium | stereo distance, parallel configuration |
| Kendo | natural scene, moving cameras, high detail | high | stereo distance, parallel configuration |
| Lovebirds1 | natural scene, natural light, high detail | medium | stereo distance, parallel configuration |
| Mobile | animation, high detail | simple | stereo distance, parallel configuration |
| Newspaper | natural scene, high detail | medium | stereo distance, parallel configuration |
| Pantomime | natural scene, medium detail | high | stereo distance, parallel configuration |

Table 2 Features of the tested sequences.

| Sequence Name | Ratio Depth/Texture in % |
|---------------|--------------------------|
| Ballet | 51.6 |
| Balloons | 28.21 |
| Book Arrival | 31.97 |
| Breakdancers | 46.27 |
| Cafe | 38.38 |
| Champagne | 52.11 |
| Kendo | 27.6 |
| Lovebirds | 23.58 |
| Mobile | 16.57 |
| Newspaper | 30.97 |
| Pantomime | 19.48 |

Table 3 Ratio between texture and depth information allowing the minimal distortion in terms of PSNR.

of depth is higher than that of the texture, the synthesis quality mostly relies on the accuracy of the depth map.

In conclusion, these results suggest that an preliminary analysis texture and depth entropies can be used as an indicator for automatic bit-rate allocation between these two types of data.

4.2 High contrast background/foreground areas

We assume that errors occurring after the synthesis process are not only more noticeable when the contrast between background objects and foreground objects is high, but also more penalized by signal-based objective metrics. To investigate this assumption, we consider the strong depth discontinuities (highlighted by an edge detection algorithm) and evaluate the standard deviation of

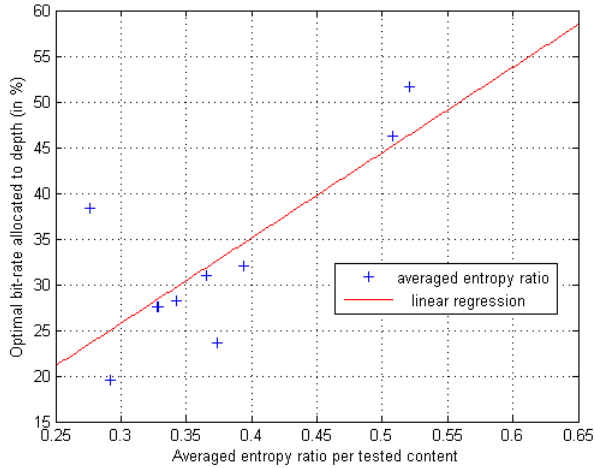


Fig. 6 Ratio of entropy between texture and depth data against optimal percentage of bit-rate allocated to depth data according to our previous experimental protocol, in terms of PSNR

the texture image around these discontinuities. Let ΔD be the gradient image of depth map D . Any pixel located at coordinates (x, y) is noted p and we consider the set of pixels Γ such as:

$$\Gamma = \{p = (x, y) | \Delta D(x, y) > 0\} \quad (2)$$

The investigated feature, denoted C , that expresses the contrast between foreground and background areas around the strong depth discontinuities, is computed as:

$$C = \frac{1}{|\Gamma|} \sum_{p=1}^{|\Gamma|} \sigma(T(p)), p \in \Gamma \quad (3)$$

where $|\Gamma|$ is the cardinality of Γ and $\sigma(T(p))$ is the standard deviation of 5×5 window centered on pixel p in texture view T . For each piece of the tested material, we compute C for left and right views and we consider the mean of these two coefficients.

Unexpectedly, the results show that the higher the contrast, the less bit-rate allocated to depth: two main point clouds are distinguishable. The point cloud corresponding to 40-55% of bit-rate allocated to depth belongs to the two toed-in camera configuration sequences. The second cloud corresponds to the parallel camera configuration sequences. So, our assumption is that despite the high contrast around objects contours, the camera configuration (and thus the distance to the virtual view) might reduce the impact of the synthesis distortions.

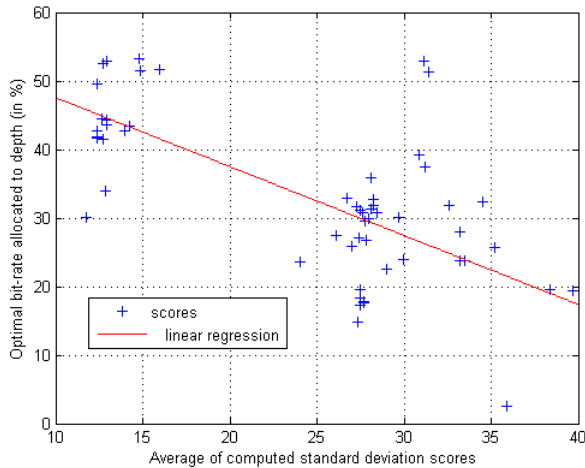


Fig. 7 Influence of high contrast background/foreground areas: Average of computed standard deviation scores around gradient pixels of depth maps against optimal percentage of bit-rate allocated to depth data according to our previous experimental protocol, in terms of PSNR

4.3 Baseline distance between cameras

This aspect is closely related to the distance of the assessed synthesized viewpoint to the reference views. We assume that there is a relationship between the structure of the scene depth and the “optimal” percentage of bit-rate allocated to depth data in MVC. According to the depth structure complexity and the baseline distance between the reference cameras, discovered areas in the novel virtual viewpoints are relatively large and difficult to fill-in by the synthesis process. Since the discovered areas are filled in with in-painting methods whose texture estimation quality differs according the used strategy, these areas are prone to perceptible synthesis errors. We aim at evaluating the influence of the discovered areas on the “optimal” percentage of bit-rate allocated to depth data. Let V_r and V_l be the original right and left view, respectively and D_r and D_l be the original right and left depth maps respectively. Let $V_{r \rightarrow v}$ the projection of V_r into the target virtual viewpoint, and $V_{l \rightarrow v}$ the projection of V_l into the target virtual viewpoint. $V_{r \rightarrow v}$ and $V_{l \rightarrow v}$ contain undetermined areas that correspond to the discovered areas. $V_{r \rightarrow v}$ and $V_{l \rightarrow v}$ are used to create logical masks $M_{r \rightarrow v}$ and $M_{l \rightarrow v}$ defined as:

$$M_{r \rightarrow v}(x, y) = \begin{cases} 0, & \text{if } V_{r \rightarrow v}(x, y) \text{ is determined} \\ 1, & \text{if } V_{r \rightarrow v}(x, y) \text{ is not determined} \end{cases} \quad (4)$$

$$M_{l \rightarrow v}(x, y) = \begin{cases} 0, & \text{if } V_{l \rightarrow v}(x, y) \text{ is determined} \\ 1, & \text{if } V_{l \rightarrow v}(x, y) \text{ is not determined} \end{cases} \quad (5)$$

Then we consider the importance I of the discovered areas according to its depth by applying the masks on the respective depth maps as follows:

$$I = \frac{1}{2 \times M \times N} \sum_{x=1}^N \sum_{y=1}^M (D_r(x, y) \times M_{r \rightarrow v}(x, y) + D_l(x, y) \times M_{l \rightarrow v}(x, y)) \quad (6)$$

where N and M are the width and height of the original image. The score I is computed for each piece of the tested material. In each case, the target virtual point is as indicated in Table 1. The results are plotted in Fig. 8. This figure shows a linear relation between the computed importance score I and the “optimal” percentage of bit-rate allocated to depth data. Although the results suggest a relationship between the discovered areas and the “optimal ratio”, the virtual viewpoint is not always known at the encoder side. This limits the use of such an indicator for automatic bit-rate control strategies.

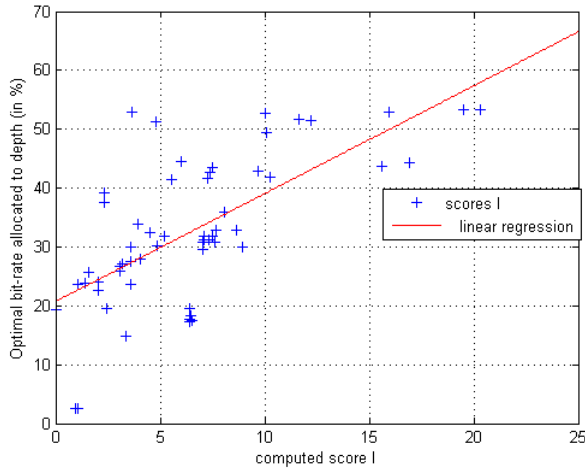


Fig. 8 Importance of discovered area against optimal percentage of bit-rate allocated to depth data according to our previous experimental protocol, in terms of PSNR

5 Conclusion

This paper aimed at determining the elements impacting on the best ratio between texture and depth data in MVD compression. Different experiments following the same basis explored several assumed factors. The experiments consisted in encoding both texture and depth data with a given compression scheme, varying the ratio between texture and depth information and analyzing the quality of the rendered virtual view. A first study consisted in investigating the influence of the encoding method on the best ratio between

texture and depth data. The test codecs were H.264/MVC coder in the first case of use, and HEVC in the second case of use. Depending on the encoding method, the attributed depth ratio allowing the best synthesized image quality in terms of PSNR was different. This suggests that bit-rate savings can be achieved if the encoding method is adapted to depth features, allowing better synthesized views quality. The experiment also showed that the ratio allocated to depth should decrease when the total bit-rate budget decreases.

Another aspect investigated in this paper concerns the impact of the input data features. A relevant remark regards the observation that the optimal ratio is significantly different depending on the sequence. The factors related to the sequences have been studied. The assumed factors influencing the best trade-off for bit-rate allocation between texture and depth data were the depth map entropy, its complexity coupled with the camera baseline distance, and the contrast of neighboring background/foreground pixel areas. The experiments revealed the existence of their impact on the best trade-off for bit-rate allocation between texture and depth data.

Despite its limitations regarding the used quality assessment tools, this study raises new topics for future work: the influence of the synthesis method on the best ratio; the investigation of the reasons why the best quality of reconstruction by using VSRS requires different depth/texture ratios depending on the content; the consideration of depth maps features for the design of depth-adapted efficient compression methods for bit-rate savings. Such investigations could be useful for the implementation of an automatic bit-rate allocation method in coding methods. In addition, this work could be extended by considering the subjective quality of the synthesized views.

References

1. A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, "Coding algorithms for 3DTV-a survey," *IEEE transactions on circuits and systems for video technology*, vol. 17, no. 11, pp. 1606–1620, 2007.
2. L. M. J. Meesters, W. A. IJsselstein, and P. J. H. Seuntjens, "Survey of perceptual quality issues in three-dimensional television systems," in *Proceedings of the SPIE*, 2003, vol. 5006, p. 313326.
3. C. Fehn et al., "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, 2004, vol. 5291, pp. 93–104.
4. I. Daribo, C. Tillier, and B. Pesquet-Popescu, "Motion vector sharing and bit-rate allocation for 3D video-plus-depth coding," *EURASIP JASP Special Issue on 3DTV*, p. 258920, 2009.
5. Emilie Bosc, Muriel Pressigout, and Luce Morin, "Focus on visual rendering quality through content-based depth map coding," in *Proceedings of Picture Coding Symposium (PCS)*, Nagoya, Japan, 2010.
6. P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proceedings of ICIP*, 2007, pp. 201–204.
7. A. Vetro, S. Yea, and A. Smolic, "Towards a 3D video format for auto-stereoscopic displays," *Proceedings of the SPIE: Applications of Digital Image Processing XXXI*, San Diego, CA, USA, 2008.

8. E. Martinian, A. Behrens, J. Xin, A. Vetro, and H. Sun, "Extensions of h. 264/AVC for multiview video compression," in *Image Processing, 2006 IEEE International Conference on*, 2006, p. 29812984.
9. E. Ekmekcioglu, S. T. Worrall, and A. M. Kondo, "Bit-rate adaptive down-sampling for the coding of multi-view video with depth information," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2008*, 2008, pp. 137–140.
10. Y. Morvan, D. Farin, and P.H.N. de With, "Joint depth/texture bit-allocation for multi-view video compression," *Picture Coding Symposium*, vol. 10, no. 1.66, pp. 4349, 2008.
11. G. Cheung, V. Velisavljevi, and A. Ortega, "On dependent bit allocation for multiview image coding with depth-image-based rendering," *IEEE Trans Image Process*, vol. 20, no. 11, pp. 3179–94, 2011.
12. K. Wegner and H. Schwarz, "Test model under consideration for hevc based 3d video coding v3.0, iso/iec jtcl/sc29/wg11 mpeg, n12744," 2012.
13. Emilie Bosc, Vincent Jantet, Muriel Pressigout, Luce Morin, and Christine Guillemot, "Bit-rate allocation for multi-view video plus depth," in *Proc. of 3DTV Conference 2011*, Turkey, 2011.
14. E. Bosc, P. Riou, M. Pressigout, and L. Morin, "Bit-rate allocation between texture and depth: influence of data sequence characteristics," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*, 2012.
15. "Jmvm software <http://iphome.hhi.de/suehring/tml/>," 2012.
16. Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model," *Signal Processing: Image Communication*, vol. 24, no. 8, pp. 666–681, Sept. 2009.
17. H. Schwarz, G. Sullivan, H. Schwarz, T. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards #x2013; including high efficiency video coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1, 2012.
18. A. Smolic, G. Tech, and H. Brust, "Report on generation of stereo video data base," Tech. Rep. Mobile 3DTV report, D2.1, 2010.