HAL
open science

Focus on visual rendering quality through content-based depth map coding

Emilie Bosc, Muriel Pressigout, Luce Morin

To cite this version:

Emilie Bosc, Muriel Pressigout, Luce Morin. Focus on visual rendering quality through content-based depth map coding. 28th Picture Coding Symposium, Dec 2010, Nagoya, Japan. hal-00541208

HAL Id: hal-00541208

https://hal.science/hal-00541208

Submitted on 30 Nov 2010

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.

FOCUS ON VISUAL RENDERING QUALITY THROUGH CONTENT-BASED DEPTH MAP CODING

Emilie Bosc, Muriel Pressigout, Luce Morin

IETR UMR CNRS 6164, Image and Remote Sensing Group/INSA of Rennes
20, avenue des Buttes de Coesmes, 35708 RENNES CEDEX 7
France

ABSTRACT

Multi-view video plus depth (MVD) data is a set of multiple sequences capturing the same scene at different viewpoints, with their associated per-pixel depth value. Overcoming this large amount of data requires an effective coding framework. Yet, a simple but essential question refers to the means assessing the proposed coding methods. While the challenge in compression is the optimization of the rate-distortion ratio, a widely used objective metric to evaluate the distortion is the Peak-Signal-to-Noise-Ratio (PSNR), because of its simplicity and mathematically easiness to deal with such purposes. This paper points out the problem of reliability, concerning this metric, when estimating 3D video codec performances. We investigated the visual performances of two methods, namely H.264/MVC and Locally Adaptive Resolution (LAR) method, by encoding depth maps and reconstructing existing views from those degraded depth images. The experiments revealed that lower coding efficiency, in terms of PSNR, does not imply a lower rendering visual quality and that LAR method preserves the depth map properties correctly.

Index Terms— 3D video coding, adaptive coding, depth coding

1. INTRODUCTION

3D video made of multiview video (MVV), or MVD is meant to provide either a relief feeling through the 3D television (3DTV) application[1] or to offer the ability to navigate into the video scene by selecting virtual viewpoints, thanks to the free viewpoint video (FVV) application.

MVD [2] refers to multiple conventional video sequences and their associated per-pixel depth signal. Due to the large amount of data to be processed in preparation for transmission, an efficient compression is essential for 3D video applications. A standard for MVV data compression, Multiview Video Coding (MVC)[3], has been addressed by the Joint Video Team (JVT) and MPEG. It relies on the extension of H.264/AVC. For MVV data, a standard exists yet, but the MPEG-3DV group is studying the case of MVD. The depth map can be considered as a monochromatic video signal and thus can be compressed by the algorithms of the state-of-the-art (MPEG-2, MPEG-4, H.264/AVC [4, 5]). Because of this property, a limited budget of the total allocated bit-rate is often thought as sufficient for depth data (5% to 10% in [6]). However, the compression of depth data can have huge consequences on the rendered views, at the decoder side: from a stereo pair of color and of depth data, a virtual intermediate viewpoint can be rendered thanks to a depth-image-based rendering (DIBR) algorithm [7, 8]; artefacts of depth images can lead to annoying visual effects on the rendered view, not to mention errors inherent in synthesis methods. As a result, measuring and optimizing the visual quality of encoded 3D video sequences can rely on the assessment of the distortion of such a rendered view. Most of the advanced 3D video coding standards refer to measurements such as PSNR as a performance estimator. And yet, as much as a good PSNR score implies that the pixel values are close from the reference's, a low PSNR score does not always indicate that the visual quality of the degraded image is as low. Furthermore, this metric is known not to be suitable for perceptual assessments, especially for 3D: first of all, PSNR score on the depth maps is not relevant to the rendering quality issue. Secondly, because of the close relations between depth information and color, the rendering quality is capital: virtual view synthesis uses depth data to realize correct color projections into the virtual viewpoint. Consequently, a distortion such as erroneous projection of a few pixels may not lead to a very low score, but it can be visually perceived as unnatural and unpleasant. On the other hand, a whole area of the image could be shifted by the compression and thus lead to a very low PSNR, while it remains visually consistent. Moreover, 3D perception is based on various depth cues, among which, occlusion, geometric perspective, etc. Artefacts occurring in those specific areas can have serious impact on the visual quality of the reconstructed view. Image quality depends on image content; yet, rate-driven coders cannot guarantee reliable identification and removal of visually irrelevant information. Previous work tried to address this issue: in [9] a scheme that preserves the discontinuities at the edges and smoothness in homogeneous

978-1-4244-7135-5/10/$26.00 ©2010 IEEE

areas uses compressed sensing. In [10], the depth map compression method is based on a multidimensional multiscale parser algorithm. It also preserves the edges.
Assuming that the more consistent the depth image is, the better the rendered view, we suggest to use a content-based compression method, so that the most relevant features of the depth map are preserved, in preparation for the rendering process. This method, namely LAR [11], was initially proposed for lossless and lossy still image compression and has been designed to ensure good visual quality.

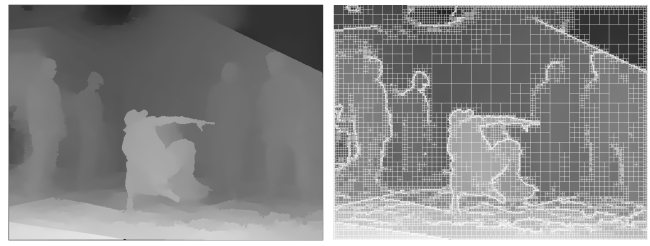
The following section discusses the H.264-based depth compression. Section 3 is devoted to LAR-based compression and its advantages for depth compression. Section 4 states the experimental protocol used to evaluate the performances of the two approaches, while section 5 discusses the results. Finally, section 6 concludes the paper.

2. H.264-BASED DEPTH CODING

The video coding standard MVC is a block oriented motion-compensation based codec developed by the Video Coding Experts Group (VCEG) and MPEG. It is an extension of the H.264/AVC standard for efficient encoding of MVV sequences. H.264/MVC improves the efficiency of intra video coding by using the spatial domain prediction. It has shown significant superior performances compared to the case where each view is independently encoded (simulcast coding). The method is based on a B-hierarchical prediction combining temporal and inter-view prediction. In our experiments, MVC is used for encoding depth information, for comparison.

3. LAR-BASED DEPTH CODING

We suggest to use a perceptually-based method in order to preserve the depth maps consistency after encoding. The LAR method is based on the assumption that an image can be considered as the combination of two components: the flat image and the local texture. The flat image models the image as a set of smooth regions separated by sharp contours, while the local texture models the details. The LAR method is a suitable candidate for depth compression because it preserves the semantic content of the image thanks to a quad-tree decomposition. The method is based on two main components: the flat codec (for the flat image) and the texture codec (for the local texture of the image). The former provides a low bit rate compressed image whereas the latter encodes the details. This basic scheme has been improved and provides many functionalities such as self-extraction of regions of interest. Considering the particular properties of depth data, we assumed that the main information of the depth map is already contained in the first component of the image, that is to say the flat image. This assumption has been confirmed by experiments not described here. Therefore, we chose to only use the flat coder for the encoding.



(a) Original depth image (b) Variable block sizes representation

Fig. 1. An example of quad-tree representation used in LAR compression method.

The flat coder is based on the principle that the local resolution can depend on the local activity. It provides a flat representation of the original image: the image is segmented into blocks of various sizes (from 16×16 to 2×2) and each block is affected the mean value of its pixels. The segmentation is driven by a quad-tree decomposition, dependant on a local gradient estimation. Consequently, small blocks of the representation are located on contours and large ones suit homogeneous areas, as illustrated on Figure 1.

The basic steps of the algorithm are described here:
-a quad-tree decomposition is computed from the original image I of size $N \times M$. The threshold value, denoted as Y, used to perform the quad-tree decomposition may affect the final representation (see Figure 1). It directly influences the detection of homogeneous areas.
-each block of the quad-tree image is associated to the final image that the flat coder outputs: each block of the flat image, defined by the quad-tree image, is affected the mean value of its corresponding pixels in the original image I.
More specifically, the compression of this representation of the image uses the Interleaved S+P algorithm and the pyramidal decomposition [12] of the image. It is useful for the prediction step. It is meant to allow lossless representation of images and scalable transmission of compressed data. The pyramid, built from I, consists of a set of images, noted as $\{Y_l\}_{l=0}^{l_{max}}$, as a multi-resolution representation of the image, where l_{max} is the top of the pyramid and $l = 0$ is the lowest level, i.e. the full resolution image. At each level, the image is expressed by:

$$\begin{cases} l = 0, & Y_0(i,j) = I(i,j), \\ l > 0, & Y_l(i,j) = \lfloor \frac{Y_{l-1}(2i,2j)+Y_{l-1}(2i+1,2j+1)}{2} \rfloor, \end{cases}$$

S+P transform is applied to each level of the pyramid, the lower level being the original image I: it is based on the 1D-S-transform applied on 2 vectors formed by 2 diagonally adjacent pixels in a 2×2 block. The couple of transformed coefficients are denoted as z_0 and z_1, and (u_0, u_1) is the diagonal couple values of a block:

$$z_0 = \lfloor (u_0 + u_1)/2 \rfloor,$$
$$z_1 = (u_1 + u_0)$$

The pyramid already stores the z_0 coefficient which is an average value of the diagonal of one 2×2 block. The term

”interleaved” of the method refers to the fact that the transformation of the second diagonal can be seen as a second S-pyramid.
From the top of the pyramid, the reconstruction of the bottom levels only require the gradient values, by prediction. From the top to the bottom, the image is reconstructed and a post-processing can be applied to the LAR flat image to remove blocking artefacts. A median filter has been used in the presented experiments.

4. PROTOCOL

The aim of these experiments was to determine the impact of compressing depth data on the visual quality of a synthesized virtual view, to study the resulting artefacts and to evaluate whether rate-distortion performances are consistent with visual rendering. So, after encoding and decoding depth data, we realized a virtual view synthesis. The virtual view synthesis is necessary in the evaluation of depth compression, because accurate depth data is essential for a good rendering quality of the synthesized view. Indeed, as illustrated in Figure 2, the method used for virtual view synthesis, consists of the reconstruction of an intermediate virtual view from depth and color data of two different viewpoints. We employed View Synthesis Reference Software (VSRS, version 3.5) [13], provided by MPEG, for view synthesis; hence, we used compressed depth maps for the virtual view synthesis. The color images remain original because our goal was to evaluate coding artefacts caused by depth compression. The experiment we held consists of encoding depth frames from two viewpoints, by either LAR compression, or H.264/MVC intra mode. We have used the very first frame of views 2, and 4 from ”*Breakdancers*” sequence (provided by Microsoft Research). Decoded depth map of views 2 and 4 were then used to synthesize view 3 and compute the PSNR score, with respect to the original color image from view 3 (see Figure 2).

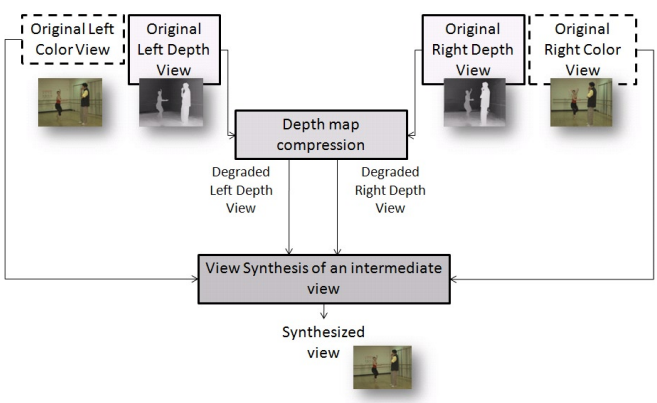
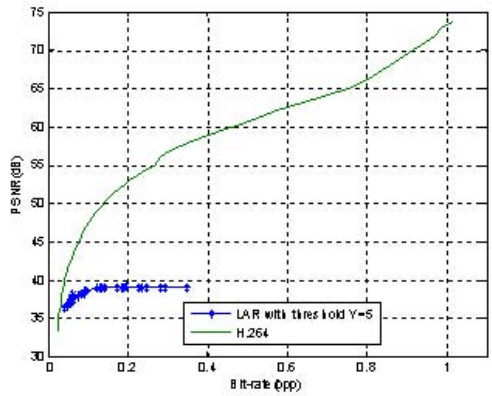


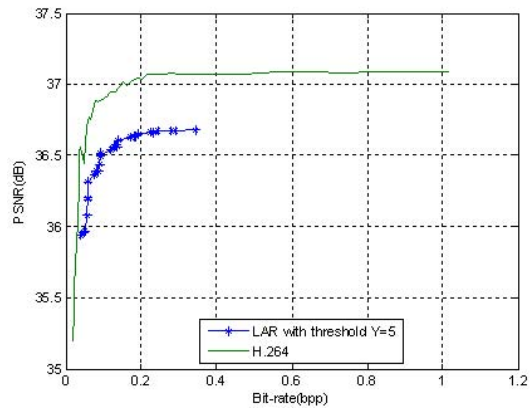
Fig. 2. Protocol.

5. RESULTS

Figure 3 and Figure 4 show the results. In Figure 3, the PSNR values are plotted over the bit-rate. According to the PSNR



(a) Distortion of Depth map



(b) Distortion of Synthesized view

Fig. 3. Rate-distortion curves for the ”*Breakdancers*” (a) depth images, and (b) synthesized images, for LAR algorithm and H.264.

scores, the H.264/AVC compression method outperforms the LAR compression method: the gap between H.264/AVC and LAR regarding the quality of depth map decoded frames is significant. However, scores differ from only 0.5dB concerning the quality evaluation of the synthesized view. This confirms our assumption that the LAR compression preserves essential depth information. Figure 4 shows a comparison of rendered views from encoded depth maps at different bit rates, with either H.264/AVC (intra-coding) or LAR. In Figure 4, the three presented areas show that although LAR provides lower PSNR than H.264/AVC intra-coding, the rendering quality evaluation suggests that LAR method achieves a rendering quality very close to H.264/AVC intra-coding or better, at the same bit-rates: Figure 4(d) shows that LAR method better preserves the vertical line, on the wall, than H.264/AVC intra-coding. In Figure 4(a) and 4(b), the visual quality of both methods is similar.

6. CONCLUSION

We investigated LAR method performances by achieving the compression of depth maps from two different viewpoints. Then the decoded depth maps were used to build up the intermediate virtual viewpoint. Results showed that H.264/AVC

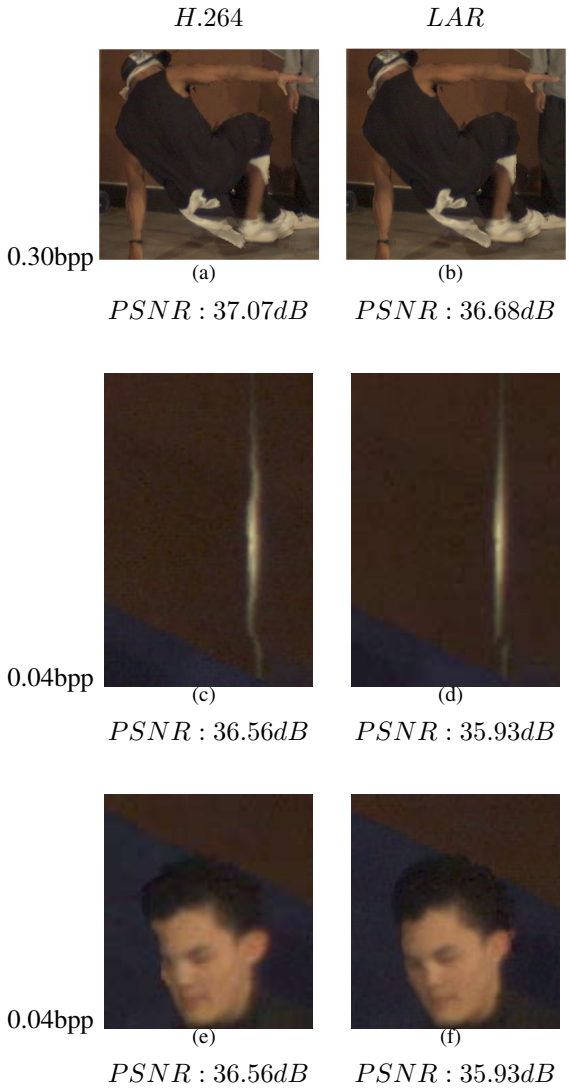


Fig. 4. Synthesized images.

compression give higher PSNR scores. However, the rendering quality evaluation showed that LAR method achieves an equal or better visual quality. Furthermore, a visualization on a stereoscopic screen gave the same results, that is to say, lower coding efficiency does not imply a lower visual quality. These results suggest that improvements can be achieved by using more perceptually driven encoding methods for 3D video. Furthermore, relations between depth and color data are still not properly understood: we haven't determined yet exactly how the bit budgets used for coding the color and depth data can affect the quality of the 3D video. We also think of improving the method by using a joint depth-color coding. In addition, the color information could be used as a tool for the post-processing filter instead of a simple median filter.

7. ACKNOWLEDGMENTS

This work is supported by the ANR-PERSEE project. We would like to acknowledge the Interactive Visual Media Group of Microsoft Research for providing the "*Breakdancers*" data set, and MPEG for the VSRS algorithm.

8. REFERENCES

[1] C. Fehn, K. Hopf, and B. Quante, "Key technologies for an advanced 3D TV system," in *Proc. of SPIE*, 2004, vol. 5599, p. 66.

[2] A. Smolic, K. Mueller, P. Merkle, N. Atzpadin, C. Fehn, M. Mueller, O. Schreer, R. Tanger, P. Kauff, and T. Wiegand, "Multi-view video plus depth (MVD) format for advanced 3D video systems," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q. 6, JVT- W*, p. 2127, 2007.

[3] T. Wiegand, G. J Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h. 264/AVC video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560576, 2003.

[4] ITU-T Recommendation H.264, "Advanced video coding for generic Audio-Visual services," 2009.

[5] G. J Sullivan, P. Topiwala, and A. Luthra, "The h. 264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," in *SPIE Conference on Applications of Digital Image Processing XXVII Paper No*, 2004, vol. 5558, p. 53.

[6] E. Martinian, A. Behrens, J. Xin, and A. Vetro, "View synthesis for multiview video compression," in *PCS*, 2006, vol. 37, p. 3839.

[7] C. Fehn et al., "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, 2004, vol. 5291, p. 93104.

[8] C. Fehn, "3D-TV using depth-image-based rendering (DIBR)," in *Proc. (PCS)*, 2004, p. 307312.

[9] M. Sarkis and K. Diepold, "Depth MAP compression VIA compressed sensing," in *Proceedings of the 16th IEEE international conference on Image processing*, 2009, p. 737740.

[10] D. B. Graziosi, N. M. M. Rodrigues, C. L. Pagliari, S. M. M. de Faria, E. A. B. da Silva, and M. B. De Carvalho, "Compressing depth maps using multiscale recurrent pattern image coding," *Electronics letters*, vol. 46, no. 5, pp. 340341, 2010.

[11] O. Deforges, M. Babel, L. Bedat, and J. Ronsin, "Color LAR codec: a color image representation and compression scheme based on local resolution adjustment and self-extracting region representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 8, pp. 974987, 2007.

[12] M. Babel, O. Deforges, and J. Ronsin, "Interleaved s+ p pyramidal decomposition with refined prediction model," in *IEEE ICIP 2005*, 2005, vol. 2.

[13] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, "Reference softwares for depth estimation and view synthesis," Apr. 2008.