Technical University of Denmark

Coding of Depth Images for 3DTV

Zamarin, Marco; Forchhammer, Søren

Publication date:
2012

Citation (APA):
Zamarin, M., & Forchhammer, S. (2012). Coding of Depth Images for 3DTV. Paper presented at 3rd Annual Workshop on Photonic Technologies and Applications : 2nd Danish-Californian Photonics Workshop, Berkeley, California, United States.

Coding of Depth Images for 3DTV

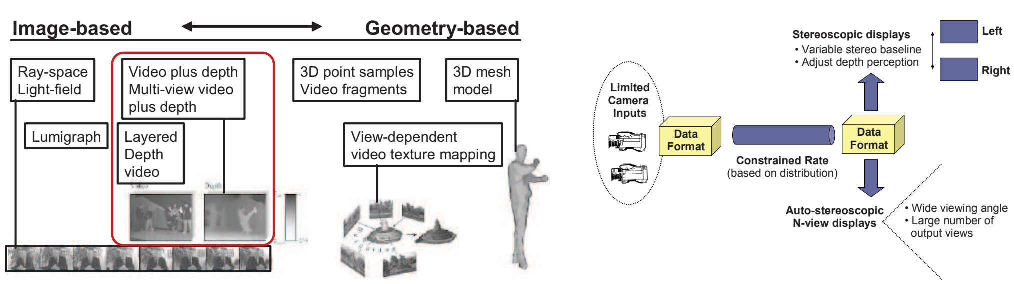
Marco Zamarin, Søren Forchhammer
Coding & Visual Communications Group, Dept. of Photonics Engineering
Technical University of Denmark, Ørsteds Plads B.343, 2800 Kgs. Lyngby, Denmark
{*mzam,sofo*}*@fotonik.dtu.dk*

Abstract: In this short paper a brief overview of the topic of coding and compression of depth images for multi-view image and video coding is provided. Depth images represent a convenient way to describe distances in the 3D scene, useful for 3D video processing purposes. Standard approaches for the compression of depth images are described and compared against some recent specialized algorithms able to achieve higher compression performances. Future research directions close the paper.

1. Introduction

A 3D scene can be represented for coding purposes through different data formats. As depicted in Fig. 1(a), these formats can be grouped into two main categories: the image-based one (e.g. light-fields and multi-view videos) in which no 3D geometry is used, and the geometry-based one (e.g. 3D mesh models) in which explicit geometric models are exploited [1]. A very popular and flexible format combining the advantages of both the categories is a hybrid one, namely “Multi-View Video Plus Depth”. This representation includes two different types of data: color and geometry. Color information is provided by a number of synchronized video streams (typically between 2 and 16) of the same scene from different viewpoints while geometric information is described by a number of depth streams. A depth image (or depth map) is a grey-scale image in which pixel values describe distances between the camera plane and the 3D points in the 3D scene.

Figure 2 shows an example of a color image with the associated depth map. Depth maps are typically used to project views to different viewpoints (“*3D warping*” [2]) in order to obtain predicted views to be used for coding purposes as done in [3], or synthesized views e.g. at the decoder side. The ISO/IEC Moving Picture Experts Groups (MPEG) is currently developing a standard for 3D video services and devices that enables advanced stereoscopic display processing and improves support for auto-stereoscopic N-view displays [4] (see Fig. 1(b)). Since they allow for high-quality view synthesis, depth images appear as a very useful side information in this scenario and have been exploited in most of the responses to the MPEG Call for Proposals on 3D Video Coding Technology [5]. Therefore,



(a) (b)

Fig. 1. (a): 3D scene representations for 3DV and FVV (reprinted from [1], copyright 2008 Springer); (b) “Target of 3D Video format illustrating limited camera inputs and constrained rate transmission according to a distribution environment. The 3DV data format aims to be capable of rendering a large number of output views for auto-stereoscopic N-view displays and support advanced stereoscopic processing”, reprinted from [4].

Fig. 2. A color image (left) and the relative depth image (right) from the “Breakdancers” data set [6].

Fig. 3. Compressed color data (left) and relative segmentation (right) used to predict shapes in the depth image.

the compression of depth images plays a crucial role in the development of a 3D video communication system and needs to be taken into account carefully in order to allow for effective performances.

2. Depth coding: standard approaches versus specialized approaches

Since depth maps are gray-scale images, they can be compressed using classical coding algorithms adopted for videos and still images - such as H.264/AVC and JPEG2000 - as done in [7] where a standard H.264/AVC Intra coder is used. However, standard approaches do not reach optimal performances since depth images have quite different characteristics than natural images: they are usually made of smooth regions divided by sharp edges. A major requirement for depth coding is the preservation of important edge information in order to allow for accurate view synthesis rendering thus avoiding sample displacement artifacts. One approach that satisfies this requirement is the Platelet-based one proposed by Morvat *et al.* [8] in which smooth regions are modeled using piecewise-linear functions and sharp edges approximated by straight lines. As shown in [9], this approach permits obtaining sharper and more accurate synthesized views compared to the cases in which H.264/AVC and MVC are used for depth compression.

Depth images are usually transmitted together with their relative color views. Thanks to the correlation between color views and corresponding depth images, color data can be effectively exploited to reduce the depth bit-rate. In [10] segmented color data are used to predict shapes in the depth image (see Fig. 3). Since within each segment depth data are expected to be very smooth, each depth segment is approximated with a parameterized plane. In case the approximation is sufficiently accurate for the target bit-rate, the surface coefficients are compressed and transmitted. Otherwise, the region is coded using a standard H.264/AVC Intra coder. Note that since the color image being segmented is the compressed one (i.e. the same one available at the decoder), no segmentation data needs to be encoded and transmitted thus permitting a considerable saving in terms of bit-rate.

Experimental results show that the proposed scheme outperforms the H.264/AVC Intra codec on depth data in terms of PSNR - with average gains of 1 to 2 dB - and edge-preservation capabilities. Moreover, this approach can outperform JPEG2000 and the Platelet-based method of [8] in terms of depth PSNR. Figure 4 compares the impact of coding artifacts on a depth image from the “ballet” dataset in case of a standard H.264/AVC Intra codec and the proposed scheme. While the H.264/AVC Intra codec introduces significant degradations of the main edges due to its

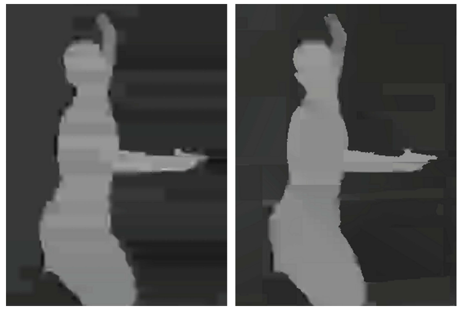


Fig. 4. Details of reconstructed depth images from the "ballet" data set [6]: standard H.264/AVC Intra (left) and plane fitting exploiting color segmentation (right), both at 0.025 bpp.

low-pass behavior, the proposed approach is able to preserve sharp edges thus permitting more accurate view synthesis operations.

3. Lossless coding of depth images

An interesting research direction considers lossless compression of depth images. Lossless coding permits avoiding rendering artifacts in synthesized views due to depth compression artifacts. In case of natural images, lossless coding typically provides compression factors of about 2 to 3, which make lossless compression not suitable for many practical applications. As for depth images, preliminary results show that much higher compression factors can be achieved when specialized algorithms are applied: 15 to 30 in case of intra coding and 20 to 50 in case of stereo depth images. Some early works like [11] show that efficient lossless compression can be obtained also in case of depth video. Such performances encourage further investigation and can motivate for lossless (and near-lossless) transmission of depth data, at least in specific scenarios.

4. Conclusion and future work

In this short paper the topic of compression of depth images and its relevance in view synthesis operations is discussed. Some standard and specialized coding approaches for depth data have been briefly presented, highlighting the advantages of the latter over the former. Future developments include depth coding using graph-based transforms, near-lossless coding of depth data for higher compression efficiency, multi-view depth image coding exploiting inter-view depth redundancy, and efficient joint color and depth coding.

References

1. A. Smolic, P. Merkle, K. Muller, C. Fehn, P. Kauff, and T. Wiegand, *Three-Dimensional Television*, ch. Compression of Multi-view Video and Associated Data, pp. 313350. Springer Heidelberg, 2008.
2. R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, second ed., 2004.
3. M. Zamarin, S. Milani, P. Zanuttigh, and G. Cortelazzo, "A novel multi-view image coding scheme based on view-warping and 3D-DCT," *Elsevier Journal of Visual Communication and Image Representation*, vol. 21, no. 5-6, pp. 462473, 2010.
4. Video and Requirements Group, "Vision on 3D Video," ISO/IEC JTC1/SC29/WG11 N10357, Lausanne, CH, Feb. 2009.
5. 3D Video AHG, "AHG Report on 3D Video Coding," ISO/IEC JTC1/SC29/WG11 m21469, Geneva, CH, Nov. 2011.
6. L. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 600608, 2004.
7. M. Zamarin, P. Zanuttigh, S. Milani, G. M. Cortelazzo, and S. Forchhammer, "A joint multi-view plus depth image coding scheme based on 3D-warping," in *Proc. of 1st ACM Int. Workshop on 3D Video Processing* (3DVP 2010), pp. 712, 2010.
8. Y. Morvan, D. Farin, and P. de With, "Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multiview images," in *Proc. of IEEE ICIP 2007*, 2007.
9. K. Müller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proc. of the IEEE*, vol. 99, pp. 643656, Apr. 2011.
10. S. Milani, P. Zanuttigh, M. Zamarin, and S. Forchhammer, "Efficient depth map compression exploiting segmented color data," in *2011 IEEE Intl Conf. Multimedia and Expo*, pp. 16, July 2011.
11. K. Kim, G. Park, and D. Suh, "Bit-plane-based lossless depthmap coding," *Optical Engineering*, vol. 49, June 2010.