



HAL
open science

Depth coding and perceived quality for 3D video

Marco Calemme, Marco Cagnazzo, Béatrice Pesquet-Popescu

► To cite this version:

Marco Calemme, Marco Cagnazzo, Béatrice Pesquet-Popescu. Depth coding and perceived quality for 3D video. QoMEX 2015, May 2015, Costa Navarino, Greece. pp.1 - 2, <10.1109/QoMEX.2015.7148116>. <hal-01423653>

HAL Id: hal-01423653

<https://hal.science/hal-01423653v1>

Submitted on 9 Jan 2017

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

Depth coding and perceived quality for 3D video

Marco Calemme, Marco Cagnazzo, *Senior Member, IEEE*, and Beatrice Pesquet-Popescu, *Fellow, IEEE*

Abstract—Multi-view video plus depth is emerging as the most flexible format for 3D video representation, as witnessed by the current standardization efforts by ISO and ITU. The depth information allows synthesizing virtual view points, and for its compression various techniques have been proposed.

We make a preliminary investigation of the effects on the synthesized views of two different approaches: object-based and block-based, from a perceptual point of view.

Index Terms—perceived quality, 3D video, multiple-views-plus-depth, contour coding, elastic curves.

I. INTRODUCTION

The video-plus-depth representation for multi-view video sequences (MVD) consists of several views of the same scene with their associated depth information, which is the distance from the camera for every point in the view [1]. Depth information allows synthesizing virtual view points, for such applications as 3D television and free-viewpoint video, but it requires ad-hoc compression technique, since those developed for texture images are not well suited for depth. As a consequence, MVD compression has attracted a huge amount of research effort in the last years, while ISO and ITU are jointly developing an MVD coding standard [2].

A key point in depth images compression is that they are not meant to be visualized but only used for rendering of virtual views. Various techniques have been specifically proposed to code the depth information associated to the views. Recent approaches include object-based coding of depths. Calemme *et al.* [3] proposed a technique using elastic deformation of curves to losslessly encode the object contours, and shape-adaptive wavelet to represent the inner parts of the objects. This techniques proved to be competitive with state-of-the-art approaches in terms of objective quality metrics (PSNR and SSIM). In this article we shall investigate the perceived video quality of synthesized views obtained with the use of depth maps coded with the technique described in [3] and we will make comparisons with HEVC/Intra.

II. CODING OF DEPTH MAPS

The depth map compression technique described in [3] is composed of two different parts: a lossless coding technique that uses elastic curves as inter-frame predictor is used for the contours of the objects; and a object-based coding technique for the interior part of the objects.

Lossless contour coding using elastic curves. Srivastava *et al.* [4] introduced a framework to model a continuous evolution of elastic deformations between two reference curves. The referred technique interpolates between shapes and makes the intermediary curves retain the global shape structure and the important local features such as corners and bends. The elastic

deformations along two given curves (Fig. 1) reproduces very well the deformations related to a change of viewpoint or a temporal evolution of an object in a sequence, given the initial and final shapes.

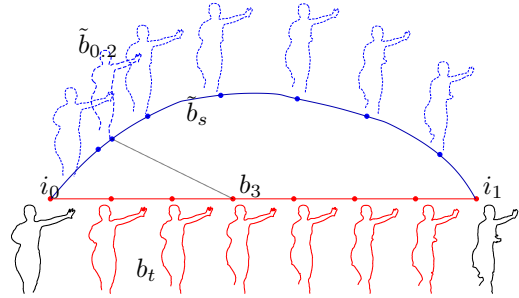


Fig. 1. Geodesic path of elastic deformations \tilde{b}_s from the curve i_0 to i_1 (in dashed blue lines). b_3 is one of the contours b_t extracted from the intermediate frames between the two reference ones, a good matching elastic curve $\tilde{b}_{0,2}$ along the path is highlighted.

Supposing that the encoder and the decoder share a representation of the initial and final shape, they can reproduce exactly the same geodesic path between them. Then, the decoder will use a suitable point of the geodesic, *i.e.* one of the dashed curves in Fig. 1, as context to encode an intermediary contour (one of the solid curves in the same figure). The encoder will only have to send a value in $[0, 1]$ to let the decoder identify this curve. The lossless coding of the contour is performed through an arithmetic coder, and the input symbol probability distribution is modified on the fly according to the elastic prediction [3].

Object-based depth maps coding technique. The lossless contour coding technique can be used to create a segmentation map of the scene with different objects and use this map in conjunction with an object-based coding technique. The proposed method [3] relies on the SA (Shape Adaptive) Wavelet Transform, followed by SA SPIHT (Set Partitioning In Hierarchical Trees), followed by an arithmetic coder for the SPIHT symbols (memoryless, without context). This provides an entirely Intra technique for the inner part of the objects.

Advantages/drawbacks of lossless coding. The validity of the object-based approach is suggested by the relevance of the contour information in the synthesis step. However, a minimum bit budget is needed for a *lossless* representation of the total rate dedicated to the depth map. The question that arises is if the benefits of a lossless contour will actually be perceived as relevant against its bit-rate cost. Therefore we propose in this work an alternative depth coding method: the

elastic prediction itself is used to generate the segmentation map to feed the object-based coder. The lack of precision of the elastic prediction is compensated by the fact that it is possible to lower the rate dramatically, at the price of a quality which can be acceptable for very low bit rate video.

III. COMPARISONS

The multiview sequences *lovebird* (provided by the ETRI/MPEG Korea Forum) and *mobile* (Philips) have been used for tests. We coded the depth maps with the technique proposed in [3], HEVC in mode Intra, and also the new method with contour prediction (instead of contour coding). We show for the two sequences the different kind of artifacts corresponding to each coding method.

Objective measurements and perceived quality. At very low bit-rates there is a non negligible difference in terms of PSNR, as shown in Fig. 2. The difference is due to the high sensitivity of the PSNR to pixel shifts, which are the result of the different geometry of the compressed scene: even if the contour is perfect, at low bit rates the inner part is very coarsely coded.

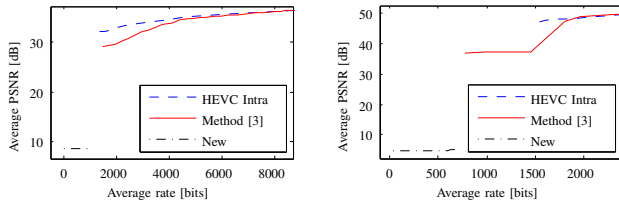


Fig. 2. PSNR comparison for *lovebird* (left) and *mobile* (right).

To make a visual comparison between the different techniques we show in Fig. 3 some image crops of the synthesized views obtained using depth maps compressed with the three techniques. HEVC/Intra is used at QP 51, the technique proposed in [3] at the lowest possible bit-rate, and the proposed technique at the same bit-rate: we want to evaluate if it is really worth to losslessly encode the contour or we can rely on the SA Coder to compensate for the imprecision. We performed a preliminary subjective test (DSIS) with 6 naive subjects and, although the differences in PSNR at low bit-rate are not negligible (see in Fig. 2), the artifacts introduced by the technique of [3] are more acceptable to the human eye: the MOS is 4.08 at the rate of 0.0019 bits per pixel, while it is 3.42 at 0.0024 bits per pixel for HEVC and 3.33 at 0.0018 bits per pixel for the new method.

In Fig. 3 the noteworthy details are highlighted with red rectangles. In fact a pixel shift is barely noticeable as we can see in Fig. 3(d1) and (d2) and (b1), which is a rare case where the predicted elastic contours are very close to the original ones. On the other hand fragmented contours, like in (b2) and (c2), or ringing effects (c1) are usually perceived as annoying. The complete test images are available at the authors' web site: <http://calemme.wp.mines-telecom.fr/?p=5>.

Conclusions. We have shown typical images for low bit-rate depth coding, casting the attention on the resulting artifacts.

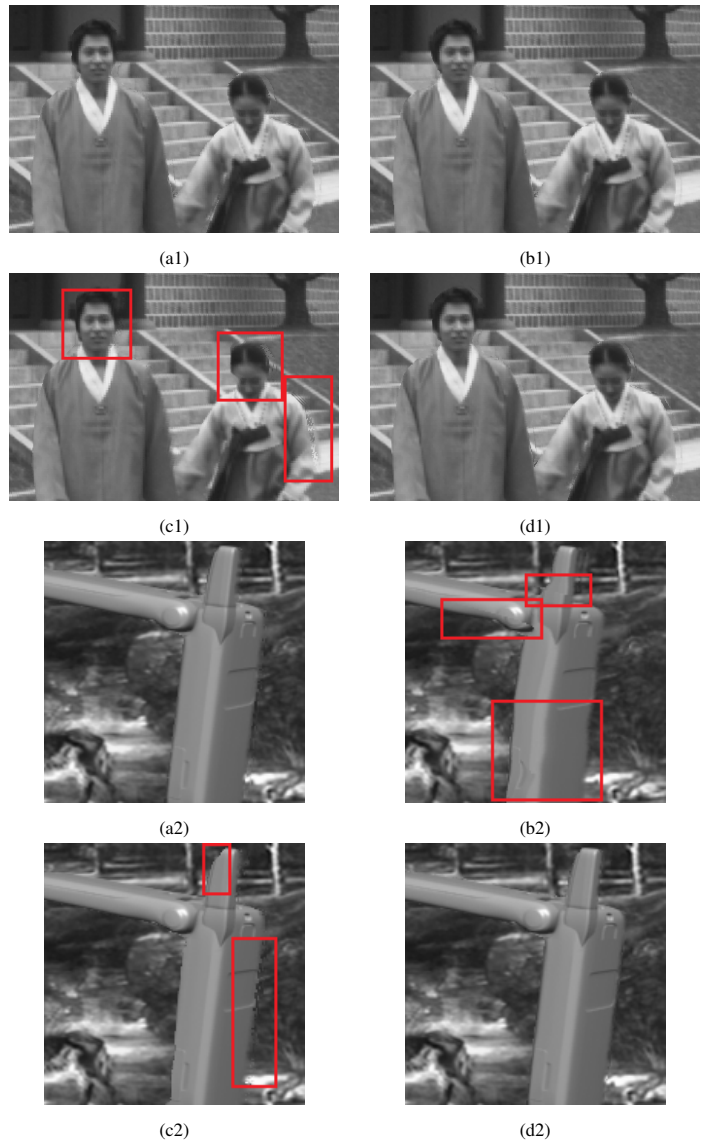


Fig. 3. Cropped detail of the synthesized view of the sequences *lovebird* (1) and *mobile* (2): (a) is the reference, obtained by the uncompressed depths; (b) is the new method; (c) is relative to the depths compressed by HEVC Intra; (d) to the method described in [3].

Given the resulting synthesized images, it seems that a contour close to the object can lead to visible benefits. However further subjective tests are needed to assess the degree of acceptance of the different techniques in various conditions.

REFERENCES

- [1] F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, Eds., *Emerging technologies for 3D video: content creation, coding, transmission and rendering*. Wiley, 2013.
- [2] "Call for proposals on 3D video coding technology," ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Tech. Rep., Mar. 2011, doc. N12036.
- [3] M. Caemmerle, M. Cagnazzo, and B. Pesquet-Popescu, "Lossless contour coding using elastic curves in multiview video plus depth," *Asia-Pacific Signal and Information Processing Association*, 2014.
- [4] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn, "Shape analysis of elastic curves in euclidean spaces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 33, no. 7, pp. 1415–1428, Sep. 2010.