PREDICTIVE DEPTH MAP CODING FOR EFFICIENT VIRTUAL VIEW SYNTHESIS

Luís F. R. Lucas[1,4], *Nuno M. M. Rodrigues*[1,2], *Carla L. Pagliari*[3],
Eduardo A. B. da Silva[4], *Sérgio M. M. de Faria*[1,2]

[1]Instituto de Telecomunicações; [2]ESTG, Instituto Politécnico de Leiria, Portugal;
[3]DEE, Instituto Militar de Engenharia; [4]PEE/COPPE/DEL/Poli, Univ. Federal do Rio de Janeiro, Brazil;
e-mails: luis.lucas,eduardo@lps.ufrj.br, nuno.rodrigues,sergio.faria@co.it.pt, carla@ime.eb.br

ABSTRACT

This paper presents a novel approach to compress depth maps envisioned for virtual view synthesis. This proposal uses a sophisticated prediction model, combining the HEVC intra prediction modes with a flexible partitioning scheme. It exhaustively evaluates the prediction modes for a large amount of block sizes, in order to find the minimum coding cost for each depth map block. Unlike HEVC, no transform is used, the residue being trivially encoded through the transmission of just its mean value.

The experimental results show that, when the encoding evaluation metric is the quality of the view synthesized using the encoded depth map against the map encoding rate, the proposed algorithm generates reconstructed depth maps that provide, for most bitrates, some of the best performances among state-of-the-art depth maps encoders. In addition, it runs approximately as fast as the HEVC HM.

Index Terms— Depth map coding, predictive coding, flexible block segmentation, depth image based rendering

1. INTRODUCTION

The recent developments in 3D displaying systems, as well as the growing consumer interest in 3D, have motivated the proliferation of 3D content and applications. The stereo-view video is the most straightforward technology, since it requires just two colour views to enable 3D perception. This technology can be viewed as a special case of the more general multiview video technology.

The multiview video format is motivated by the recent autostereoscopic displays based on a large number of views. By presenting views for different observation view-points, these displays provide a more realistic 3D perception. Since the multiview video format uses a large number of views, efficient compression techniques become a mandatory requirement. In this sense, the Multiview Video Coding (MVC) standard has been proposed, as an extension of the state-of-the-art single view video encoder H.264/AVC [1]. MVC exploits inter-view redundancy in order to better encode the views; however, a high bitrate is usually required.

The video+depth format [2] has been proposed by the MPEG group as an alternative to MVC. The main idea is to allow the synthesis of an arbitrary number of intermediate views, through a process denominated depth image based rendering (DIBR), in which only a small set of views and depth maps are encoded. Besides the significant bitrate savings in the multiview representation, the video+depth format also provides backward compatibility with 2D systems. Furthermore, it constitutes a compatible representation format across different multiview displays, since the number of generated views may be variable.

Unlike the colour views, the information in depth maps is not directly presented to the users. Depth maps represent the scene geometry based on the distance of the pixels in each object to the cameras. They are mostly constituted by planar areas delimited by sharp edges on object boundaries. Since depth is intended to be used by DIBR, proper compression of depth maps is important.

An immediate approach for depth map coding is to use traditional encoders. Most of them (e.g., JPEG2000, H.264/AVC and HEVC) use transform-based residue coding, such as the Discrete Cosine Transform (DCT) or the Wavelet Transform, as still images and video tend to have irrelevant information at higher frequencies. Although transform-based encoders provide backward compatibility with the existing technology, they usually cause some coding artifacts on depth maps around sharp edges, mainly in the presence of high frequencies. These artifacts are undesirable as they degrade the performance of the rendering process.

In order to minimize the coding artifacts on depth maps, alternative methods for efficient depth map coding have been proposed in literature. Some of the most efficient proposals, in a rate-distortion sense, present a high computational complexity. This is the case of the methods presented in [3] and [4], that encode depth maps based on linear piecewise functions, as well as of the Multidimensional Multiscale Parser [10], originally developed for still image coding, which is based on the pattern matching paradigm. The algorithm in [5] combines graph-based transform (GBT) and transform domain sparsification (TDS) in one unified optimization framework. The idea is to select the simplest transform per block, that leads to the sparsest representation, resulting in an efficient compression. The main drawbacks of this approach are the coding overhead to specify the edges and the complexity associated to the transform domain sparsification. In [6], a mesh-based depth map coding is proposed using an adaptive binary triangular tree. One disadvantage of this method is the large number of triangular patches placed along the

This project was funded by FCT - "Fundação para a Ciência e Tecnologia", Portugal, under the grant SFRH/BD/79553/2011. The third author acknowledges the financial support of CAPES and CNPq.

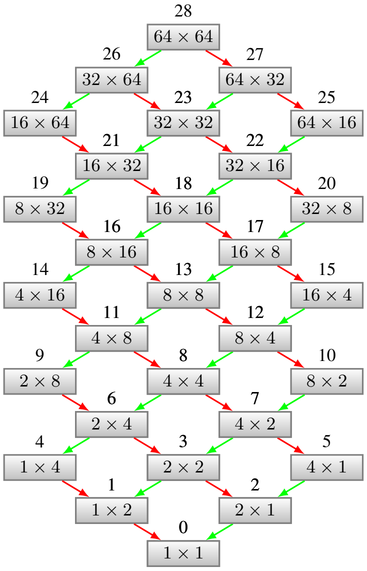


Fig. 1. Possible block scales in PDC.

edges, due to the use of a regular grid.

In this paper, an alternative depth map coding algorithm is investigated that generally outperforms the proposals in the literature, while maintaining an affordable computational complexity. This proposal is based on the flexible partitioning used by the algorithm presented in [3], combined with an improved predictive framework and a new residue encoding scheme. As a result, the computational complexity has been reduced up to 50 times relative to the one of [3], so that the proposed method runs as fast as the High Efficiency Video Coding (HEVC) algorithm [7]. In terms of the synthesized view quality against the depth map bitrate, the proposed algorithm outperforms state-of-the-art depth map coding proposals, while keeping the computational complexity within reasonable limits.

This paper is organized as follows. Section 2 describes the proposed algorithm for depth map coding. Section 3 present and discusses the experimental results, and the conclusions are presented in Section 4.

2. PREDICTIVE DEPTH CODING

The proposed Predictive Depth Coding algorithm (PDC) relies on a sophisticated predictive framework and a flexible block segmentation scheme, used within an efficient rate-distortion optimization loop. In this paper we have restricted our investigation to a version of PDC that relies only on intra coding techniques, and therefore only encodes individual depth images. The following sub-sections detail the main parts of the algorithm: block segmentation scheme, predictive framework, residue coding and rate-distortion optimization.

2.1. Block segmentation scheme

PDC uses a block segmentation tree inspired on the one used in [3], with the maximum block size increased to 64 × 64 pixels. Larger block sizes are advantageous at lower bitrates, as they allow to encode more pixels per coding unit.

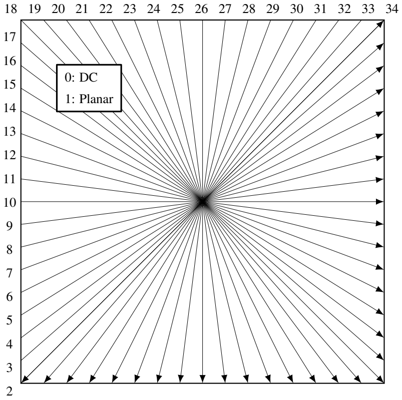


Fig. 2. Set of 35 prediction modes used in the proposed depth map coding algorithm.

The algorithm also uses a flexible block segmentation scheme, where each block may be recursively segmented either horizontally or vertically. Figure 1 presents all the 29 possible block scales in PDC.

2.2. Predictive framework

The work in [3] has shown that the prediction model has an important role in depth map coding. This question was further investigated and it was concluded that depth maps may be significantly better compressed by using more sophisticated predictive schemes. In fact, the prediction model constitutes the most important tool of the proposed encoder. Combined with the flexible block segmentation scheme in Figure 1, this framework is a good prediction scheme for depth maps. The predictive framework uses a set of 35 modes as in HEVC (see Figure 2). By using this sophisticated predictive framework across such a wide range of block scales, the PDC algorithm is able to efficiently represent edges and smooth regions of depth maps. Edges are well predicted by the directional modes, while the DC and planar modes are suitable to predict the smooth areas.

Given the large amount of block scales in PDC, some of the directional modes in Figure 2 may be irrelevant at certain scales. For example, narrow blocks may not take advantage of all the directional prediction modes, as some of them may show similar prediction results. Therefore, in order to improve the modes entropy coding and to reduce the computational complexity of the algorithm, some prediction modes at specific block scales were disabled. Table 1 presents the available prediction modes for the different block scales in the algorithm.

2.3. Residue coding

Given the high efficiency of PDC's prediction, most samples of its residue are zero. However, in order to better represent some blocks where the residue is nonzero, we transmit the mean value of the residue, which generates a small number of bits and has negligible impact on the computational complexity. The residue mean value is quantized using a piecewise-uniform quantizer. The values of the quantization step size for the intervals $[0...10]$, $]10...22]$, $]22...86]$ and $]86...255]$, are, respectively, 1, 4, 8 and 13. The PDC algorithm also allows to transmit one flag indicating that the residue is null.

Table 1. Prediction modes available at each block scale.

Block size $m \times n$	**Used directional modes**
$m \geq 8$ and $n \geq 8$	all modes
$m \geq 8$ and $n = 4$	all except 19, 21, 23, 25, 27, 29, 31, 33
$m = 4$ and $n \geq 8$	all except 3, 5, 7, 9, 11, 13, 15, 17
$m = 4$ and $n = 4$	even modes
$m < 4$ and $n \geq 4$	modes 2, 10, 18, 26, 34
$m \geq 4$ and $n < 4$	modes 2, 10, 18, 26, 34
$m < 4$ and $n < 4$	no mode

Note that in [3] a much more complex method for residue coding was used, consisting of linear fitting together with dictionary-based coding.

2.4. RD optimization procedure

The rate-distortion optimization procedure has an important role in the coding performance and computational complexity of the proposed PDC algorithm. In order to get the best coding performance, the coding cost of each 64×64 block of the depth map is minimized.

First, the fully expanded coding tree representing all the possible coding decisions is generated. This procedure implies two optimization functions. The first one tests all of the prediction modes in each sub-block that is generated by segmenting one block down to a minimum scale where the prediction can be applied. A fast decision for the prediction is used by choosing the mode that generates the minimum residue distortion. The second optimization function is applied over the residue block that may be further segmented and approximated using the mean value of each of its sub-blocks. Thus, a residue coding tree is generated for each predicted sub-block.

Then, the optimal segmentation tree is determined by applying a bottom-up pruning technique over the fully expanded tree. The optimal segmentation tree $\mathcal{T}$ is minimized according to the following Lagrangian cost function:

$$J(\mathcal{T}) = D(\mathcal{T}) + \lambda R(\mathcal{T}), \qquad (1)$$

where $D(\mathcal{T})$ is the block distortion and $R(\mathcal{T})$ is the bitrate required to encode the optimal tree $\mathcal{T}$. During the pruning procedure, each node of the coding tree is evaluated, in order to optimize the overall cost of the coding tree. Whenever the cost of a parent node is larger than the sum of the costs of its children nodes, the block remains partitioned.

Traditional image coding standards use the mean-squared error (MSE) as distortion criterion. However, since depth maps mainly consist of smooth areas bounded by sharp transitions, a different distortion measure may be used. Our experiments have shown that the mean absolute error (MAE) is preferable, as the compressed depth maps optimized according to MAE tend to generate better rendered images.

The rate function counts the number of bits required to encode the symbols using an adaptive arithmetic encoder. The symbols to be encoded are the residue mean value and prediction mode, as well as additional symbols to indicate the block segmentation mode (segmented/non segmented and segmentation direction).

Although the PDC algorithm uses a trivial residue coding scheme, its rate-distortion optimization process is exhaustive and the test of 35 prediction modes for all block segmentation options is time consuming. In this context, we used some sub-optimal conditions that restrict the expansion of the fully expanded tree. During the prediction optimization, whenever the distortion represents less than 25% of the total Lagrangian cost of the block (given by equation 1), the segmentation of the block is stopped. In the residue optimization process, a block is not segmented when the Lagrangian cost of one child block is larger than the cost of its parent.

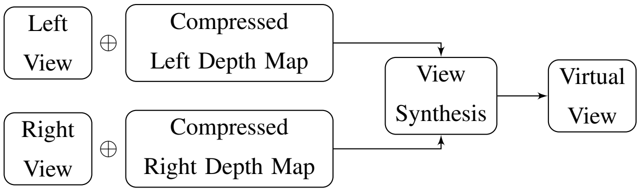


Fig. 3. Experimental framework.

The computational complexity of the algorithm was also reduced by combining two fast coding possibilities. The first one is mainly used at low bitrates. It consists of applying prediction only for scales larger than or equal to 18 (blocks larger than or equal to 16 $\times$ 16 pixels, as shown in Figure 1). The other approach is frequently used at high bitrates. It consists of dividing the 64 $\times$ 64 block into 16 sub-blocks of size 16 $\times$ 16 that are independently optimized and encoded. In this approach the minimum scale where prediction may be applied is 8 (block size 4 $\times$ 4). Both coding approaches are tested for each 64 $\times$ 64 block and the one that presents the minimum coding cost (equation 1) is chosen.

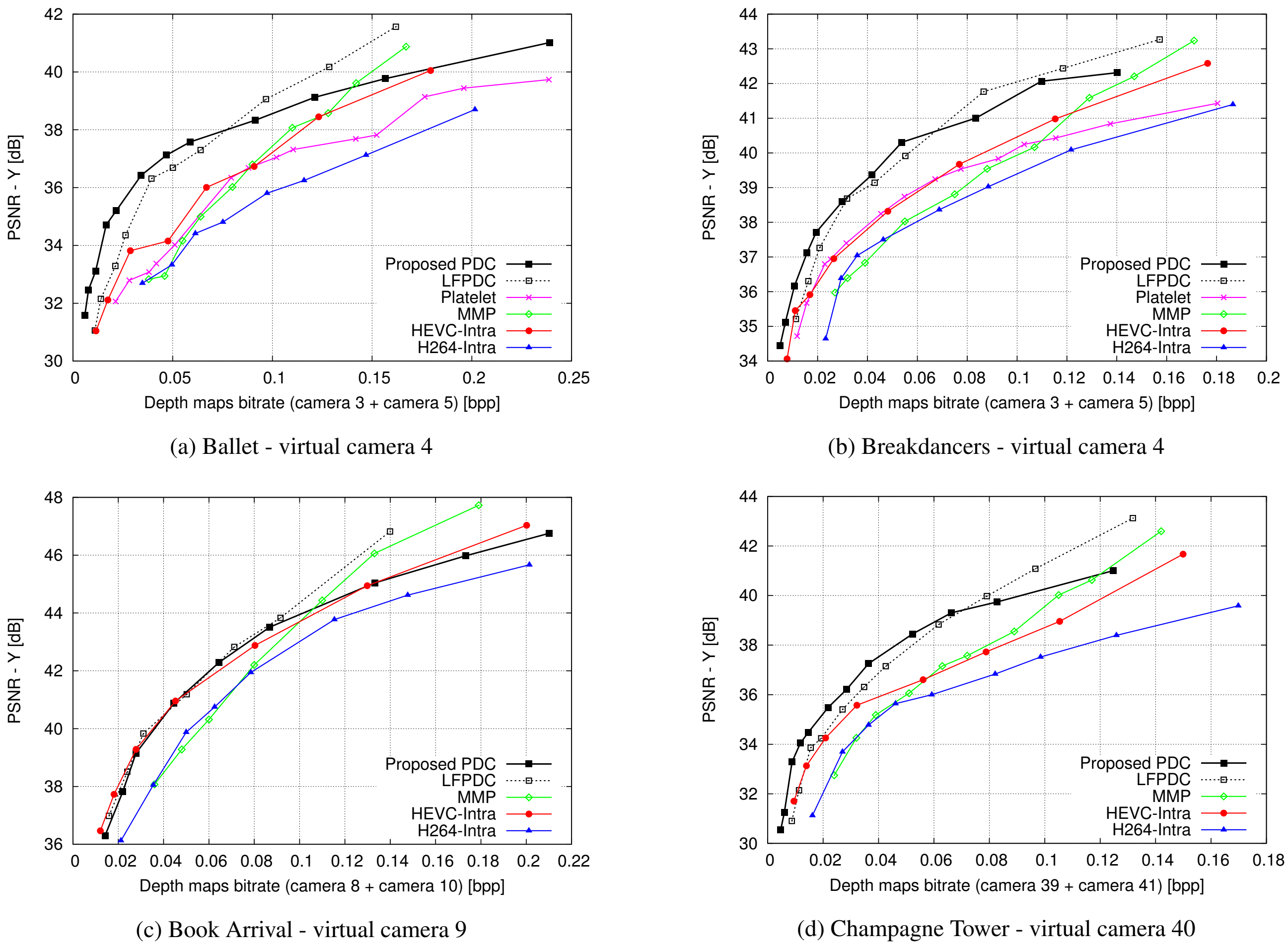
3. EXPERIMENTAL RESULTS

The performance of the proposed PDC algorithm was evaluated by analyzing the quality of the synthesized views using compressed depth maps. Its source code can be found in [8]. We have used the first frame of four test sequences, *Ballet* (camera 4), *Breakdancers* (camera 4), *Book Arrival* (camera 9) and *Champagne Tower* (camera 40) in these experiments. The first two sequences are available at *http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload*, while the others are courtesy of *FHG-HHI* and *Tanimoto Lab (Nagoya University)*, respectively.

In order to synthesize the virtual view associated to camera n, the compressed depth maps associated to cameras $n-1$ and $n+1$ (left and right cameras) were used together with the original views (luminance signal) associated to the same cameras. The software VSRS-3.5 [9] was used to perform the DIBR process. Figure 3 illustrates the framework employed.

The PDC performance was compared against that of other algorithms for which there was access to source code or experimental results. Depth maps were encoded using the transform-based standards HEVC (software HM-9.1rc1, intra main configuration) [7], and H.264/AVC (software JM-18.0 using the intra profile at level 4.0) [1]. The results for the Platelet algorithm [4] are only presented for the sequences *Ballet* and *Breakdancers*, as available at *http://vca.ele.tue.nl/demos/mvc/PlateletDepthCoding.tgz*. The performance of the depth maps encoded using MMP [10] and the method in [3] (denoted here as LFPDC, meaning Linear Fitting and Predictive Depth Coding) are also presented.

Figure 4 illustrates the rate-distortion performances of the mentioned methods for a virtual view of the first frame of each test se-



(a) Ballet - virtual camera 4

(b) Breakdancers - virtual camera 4

(c) Book Arrival - virtual camera 9

(d) Champagne Tower - virtual camera 40

Fig. 4. PSNR results of the synthesized views using the compressed depth maps together with original views.

quence. The objective qualities of the synthesized views are shown in terms of the PSNR in relation to a reference virtual view synthesized using the original (uncompressed) depth map. The shown bitrate corresponds to the sum of the rate used by both left and right compressed depth maps.

The results show that the proposed method achieves the best rate-distortion performances for most bitrates and sequences. When compared to the transform-based standards and the Platelet algorithm, PDC achieves the best results for all the bitrates, although for the *Book Arrival* sequence its performance is very similar to the one of HEVC. When compared to MMP and LFPDC, one may observe the advange of PDC at low rates. At very high bitrates the new method may present some rate-distortion performance loss in relation to both MMP and LFPDC. This is a consequence of the suboptimal conditions implemented to reduce the computational complexity described in Subsection 2.4. However, this usually occurs above 40 dB, when the difference to the reference view is nearly imperceptible. In addition, this rate-distortion performance loss should be weighted by the fact that PDC runs more than 50 times faster than MMP and LFPDC, and still presents better rate-distortion performance at low and medium bitrates.

HEVC minimizes a cost funcion using an MSE-based distortion and PDC uses the MAE as distortion. Therefore, one could argue that PDC's performance gain over HEVC could come mainly from the use of the MAE distortion criterion, and MAE-based HEVC might have the same performance gains over HEVC. In order to clarify this, we have performed experiments in which HEVC has been modified to use the MAE distortion metric instead of the MSE. The experiments, not shown here due to space limitations, show that the MAE-based HEVC performs worse than HEVC for depth maps. This justifies the investigation of the PDC algorithm for depth map encoding.

4. CONCLUSIONS

This paper presents an alternative depth map coding algorithm that relies on a sophisticated predictive model and flexible block segmentation scheme, together with an MAE distortion criterion. Since the algorithm strongly relies on prediction, a trivial residue coding method based on the mean value of the residue block has been used. Experimental results show that the compressed depth maps using the proposed PDC algorithm achieve state-of-art-results, while maintaining a level of computational complexity similar to that of the HEVC.

5. REFERENCES

[1] ITU-T and ISO/IEC JTC1, "Advanced video coding for generic audiovisual services," *ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC)*, 2010.

[2] Philips Applied Technologies, "MPEG-C part 3: Enabling the introduction of video plus depth contents," 2008, Suresnes, France.

[3] L.F.R. Lucas, N.M.M. Rodrigues, C.L. Pagliari, E.A.B. da Silva, and S.M.M. de Faria, "Efficient depth map coding using linear residue approximation and a flexible prediction framework," *to appear in IEEE Int. Conf. on Image Proc.*, September 2012.

[4] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Müller, P. H. N. de With, and T. Wiegand, "The effects of multiview depth video compression on multiview rendering," *Image Communications*, vol. 24, pp. 73–88, Jan. 2009.

[5] G. Cheung, Woo-Shik Kim, A. Ortega, J. Ishida, and A. Kubota, "Depth map coding using graph based transform and transform domain sparsification," pp. 1 –6, oct. 2011.

[6] M. Sarkis, W. Zia, and K. Diepold, "Fast depth map compression and meshing with compressed tritree," vol. 5995, pp. 44–55, 2010.

[7] G.J. Han, J.R. Ohm, Woo-Jin Han, Woo-Jin Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1, 2012.

[8] http://www.lps.ufrj.br/profs/eduardo/pdc.

[9] M. Tanimoto, T. Fujii, and K. Suzuki, "View synthesis algorithm in view synthesis reference software 3.5 (VSRS3.5) Document M16090, ISO/IEC JTC1/SC29/WG11 (MPEG)," May 2009.

[10] D.B. Graziosi, N.M.M. Rodrigues, C.L. Pagliari, E.A.B. da Silva, de S.M.M. Faria, M.M. Perez, and M.B. de Carvalho, "Multiscale recurrent pattern matching approach for depth map coding," pp. 294–297, Dec. 2010.