



EFFICIENT DEPTH MAP COMPRESSION EXPLOITING CORRELATION WITH TEXTURE DATA IN MULTIREOLUTION PREDICTIVE IMAGE CODERS

Khouloud Samrouth, Olivier Deforges, Yi Liu, François Pasteau, Mohamad
Khalil, Wassim Falou

► To cite this version:

Khouloud Samrouth, Olivier Deforges, Yi Liu, François Pasteau, Mohamad Khalil, et al.. EFFICIENT DEPTH MAP COMPRESSION EXPLOITING CORRELATION WITH TEXTURE DATA IN MULTIREOLUTION PREDICTIVE IMAGE CODERS. 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Hot topics in 3D coding., Jul 2013, San Jose, United States. <hal-00881063>

HAL Id: hal-00881063

<http://hal.univ-nantes.fr/hal-00881063>

Submitted on 7 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.

EFFICIENT DEPTH MAP COMPRESSION EXPLOITING CORRELATION WITH TEXTURE DATA IN MULTIREOLUTION PREDICTIVE IMAGE CODERS

SAMROUTH Khouloud^{1,2}, DEFORGES Olivier¹, LIU Yi¹, PASTEAU François¹, KHALIL Mohamad²,
FALOU Wassim²

¹ IETR - Image group lab, CNRS UMR 6164, INSA de Rennes

² LaSTRe, EDST, Lebanese University

{khouloud.samrouth1,olivier.deforges,yi.liu,francois.pasteau}@insa-rennes.fr,
{mohamad.khalil,wafalou}@ul.edu.lb

ABSTRACT

New 3D applications such as 3DTV and FVV require not only a large amount of data, but also high-quality visual rendering. Based on one or several depth maps, intermediate views can be synthesized using a depth image-based rendering technique. Many compression schemes have been proposed for texture-plus-depth data, but the exploitation of the correlation between the two representations in enhancing compression performances is still an open research issue. In this paper, we present a novel compression scheme that aims at improving the depth coding using a joint depth/texture coding scheme. This method is an extension of the LAR (Locally Adaptive Resolution) codec, initially designed for 2D images. The LAR coding framework provides a lot of functionalities such as lossy/lossless compression, low complexity, resolution and quality scalability and quality control. Experimental results address both lossless and lossy compression aspects, considering some state of the art techniques in the two domains (JPEGLS, JPEGXR). Subjective results on the intermediate view synthesis after depth map coding show that the proposed method significantly improves the visual quality.

Index Terms— 3D, depth, texture, prediction, joint coding

1. INTRODUCTION

New highly advanced multimedia video systems, such as 3-D television (3DTV) and free-viewpoint TV, offer more realistic 3-D scenes and allow the user to freely navigate within real-world visual scenes by transmitting video and the corresponding depth data for several viewpoints and synthesizing intermediate viewpoints [1, 2].

Such 3-D autostereoscopic systems rely on depth information to render new views [3, 4]. That is why recent research [5–7], highlights the influence of depth-image compression and its implications on the quality of video-plus-depth virtual view rendering. The state of the art compression schemes such as H.264/AVC, its extension H.264/MVC

and the emerging video coding standard, HEVC, provide efficient solutions for image/video compression. However, they are not specially designed for depth compression and 3D extensions are still in progress.

In this context, new research has been devoted to the enhancement of the coding performance over depth data in order to save depth-coded bit-rate and to enhance the visual quality of synthesized views [8–11]. We categorize these into two groups: depth independent component approaches [8, 9] and texture/depth correlation approaches [10, 11]. The first group deals with the depth map as an independent component without taking advantage of the correlation between texture and depth data. The other group intends to compress the depth component exploiting its correlation with texture view.

While many different solutions have been proposed for the compression of depth data, few exploit the correlation between texture and depth and are often not taking into account the visual quality of rendered virtual views. Such a correlation concept has been proposed in the literature for natural images, where correlation was applied between different color components. Such approaches did not give satisfying results [12]. In this paper, we propose a texture/depth correlation-based compression method for predictive image coders. We apply this method on LAR (Locally Adaptive Resolution). LAR is a global coding framework providing a lot of functionalities such as **unique codec for lossy and lossless compression**, resolution and quality scalability, partial/full encryption, Region of Interest coding, Rate Control and RDO [13]. In this paper, we upgrade the LAR from a 2D to a 3D coding framework. We propose a scalable joint depth/texture coding scheme. It consists firstly in coding the texture at low resolution. Then, this low resolution texture is used to enhance the depth coding. Finally, the texture is refined to obtain the high resolution texture. So next, we will focus on the enhancement depth coding using texture images. Experiments covered objective performances for lossy/lossless coding as well as subjective performances.

The remainder of this paper is organized as follows: in Section 2, the LAR codec framework is described in detail.

In Section 3 and Section 4, the proposed depth image coding method is explained. Section 5 provides experimental results for the coding performance and the subjective quality test of synthesized intermediate views. Finally, we conclude this paper in Section 6.

2. LAR CODER FRAMEWORK

Locally Adaptive Resolution (LAR) [14] is an efficient content-based 2D image coder for both lossless and lossy image compressions. LAR relies on a local analysis of image activity that leads to a quadtree, a non-uniform block representation. The common version of LAR supports two coding layers given in Fig. 1. The first layer, the Flat coder, is based on the quadtree partitioning and provides global information in various sized blocks enabling low bit-rate while preserving contours. The second layer, the Local Texture coder, holds local information details by refining the non decomposed blocks during the first layer.

Both layers adopt a multi-resolution coding scheme. It consists of a dyadic pyramidal decomposition on L levels of the original image. Then, the interleaved S+P scheme [15] was introduced to enable efficient compression in both lossy and lossless modes. The S+P transform (S-Transform + Prediction) is based on the S-Transform applied on the 2 vectors formed by 2 diagonally adjacent pixels in a 2×2 block as shown in Fig. 3. Let (u_0, u_1) be the pixels of the first diagonal; (u_2, u_3) pixels of the second diagonal, and (y_0, y_1) denote the S-Transform coefficients, where y_0 and y_1 represent the mean and the gradient of a diagonal of a 2×2 block respectively.

$$\begin{aligned} y_0^1 &= \lfloor (u_0 + u_1)/2 \rfloor, & y_0^3 &= \lfloor (u_2 + u_3)/2 \rfloor. \\ y_1^2 &= u_1 - u_0, & y_1^3 &= u_3 - u_2. \end{aligned} \quad (1)$$

When using the Interleaved S+P scheme within the pyramidal decomposition, y_0^1 coefficients of a given level are automatically retrieved from the upper level. Then, only three coefficients y_1^2 , y_0^3 and y_1^3 have to be estimated for each level. Therefore, S transform is followed by two gradients and one mean value prediction between pyramid levels [15]. The prediction errors are quantized and then coded by an entropy coder. A bi-linear filter is used to smooth block effects as a fast post-processing technique.

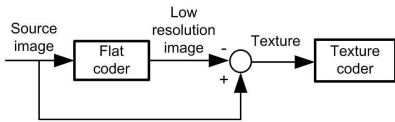


Fig. 1. General scheme of two-layer LAR coder

An important feature of LAR is that the number of operations to process is proportional to the number of blocks in the QuadTree. It means that LAR provides a scalable complexity

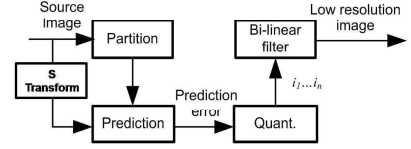


Fig. 2. Flat coder scheme at a given level of the pyramid

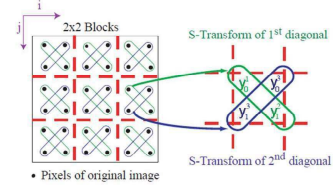


Fig. 3. S transform

at both the coder and the decoder sides. At low bit-rates, the QuadTree is composed of only a few blocks, enabling a fast coding and decoding process. For depth images, we only code with the first layer of the LAR codec (see Fig. 2). Therefore, we have two parameters to handle: the homogeneity threshold parameter Th for the quadtree decomposition and the quantization parameter Q . When processing pyramidal levels for different resolutions, Q is multiplied by the F_{Li} factor to obtain the quantization factor Q_{Li} for level Li (eq.2). Thus, the Flat LAR encodes images in different qualities by using a combination between threshold parameter Th and quantization parameter Q : lossless coding with $Th = 0$ and $Q = 1$ (full decomposition); lossy coding with other parametric combinations.

$$Q_{Li} = Q \cdot F_{Li}, \text{ for level } Li \quad (2)$$

To facilitate the parametrization, it is generally preferable to adopt just one parameter. In order to find the optimal pair $\{Q; Th\}$, considering the special characteristics of depth images that differ from natural textured images [16], a large set of experiments on 40 images has been performed. We found that $Th = \frac{1}{3}Q$ and $Th = \frac{2}{3}Q$ generally give the optimal results for encoding depth maps.

3. PROPOSED GLOBAL CODING SCHEME

The proposed global coding scheme is given in Fig. 4. In the first step, we apply the quadtree partitioning on the depth image for a given Th . The resulting grid is called Grid_Depth. Then, we use this Grid_Depth to encode the texture image at a low bit-rate. We obtain a low resolution block representation of the texture image, called Low Resolution Texture. Next, a joint depth/texture coding is used to encode the depth image. This joint coding will be detailed in Section 4. Finally, in the second step, Grid_Depth is refined using the original texture image to obtain a more detailed grid, called Grid_Texture.

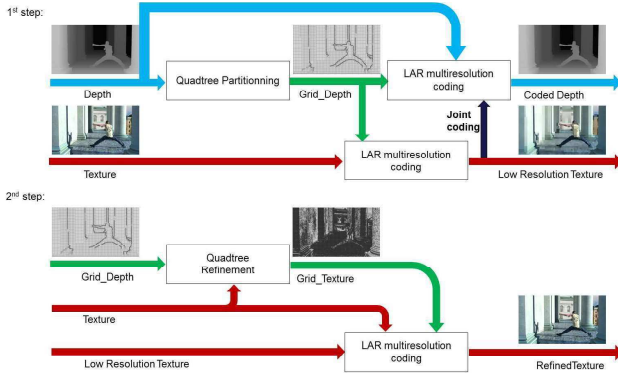


Fig. 4. Global coding scheme

Therefore, this Grid_Texture is used to refine the texture image. Consequently, the Low Resolution Texture and the Refined Texture offer scalability in quality for the texture image.

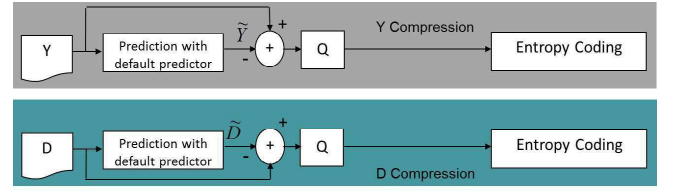
In this paper, we only focus on the depth image coding. In the next Section, we introduce the proposed depth/texture joint coding algorithm that improves the depth coding performances.

4. IMPROVED DEPTH PREDICTION

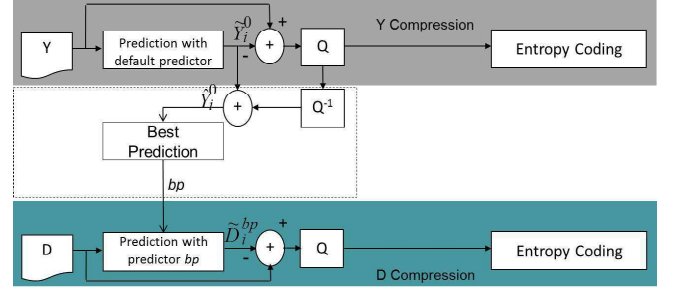
In this Section we present an approach for depth coding exploiting the correlation between depth and texture representations. This approach aims at improving the prediction of the depth component D using the best predictor of previously encoded Luma component Y of the corresponding texture image. Such a process already exists in the literature, for example in JPEGLS [17] but only between color components (R,G,B). We introduce the following notations.

- Y_i : gradient value of luma component at pixel i in texture image. Y_i can represents respectively the y_1^2 or y_1^3 shown in figure 3.
- \tilde{Y}_i^j : prediction of Y_i through predictor j , $j \in [0, \dots, \text{NbPredictors}]$.
- \hat{Y}_i^j : reconstruction of Y_i through predictor j .
- $\text{prediction}(Y_i, j)$: process returning \hat{Y}_i^j .

Fig. 5.a shows the classic coding scheme. The depth component is coded independently of the texture component Y . The proposed joint coding scheme is shown in Fig. 5.b. It was inspired from the work done by Pasteau et al in [12]. We apply the Best-Prediction algorithm on the depth component instead of color components. After predicting the Luma component Y using the default predictor (predictor 0), different predictors are reapplied on Y . Then, the best predictor can be selected by minimizing the distance between $\hat{Y}_i^j - \tilde{Y}_i^0$ according to the norm 1. Therefore, this best predictor, called



(a) Classic coding scheme



(b) Proposed coding scheme

Fig. 5. Classic coding scheme vs joint coding scheme

bp , can be used on the depth component D to obtain a better prediction. Figure 5.b presents the implementation of the Best-Prediction procedure in the overall coding scheme and procedure 1 explains the search for the best predictor.

The the Best-Prediction process, executed at the coder, is bit-rate cost free as it can be performed at the decoder without transmitting any additional information.

Procedure 1: Best-Prediction

Input: $Y_i; \hat{Y}_i^0$
Output: bp
 $best = +\infty, bp = 0$
for all j **do**
 $\tilde{Y}_i^j = \text{prediction}(Y_i, j)$
 if $|\tilde{Y}_i^j - \hat{Y}_i^0| < best$ **then**
 $best = |\tilde{Y}_i^j - \hat{Y}_i^0|$
 $bp = j$
 end if
end for
{return the best predictor according to Y }
return bp

5. EXPERIMENTAL RESULTS

We tested the proposed scheme on a large set of MPEG 3D reference sequences (*Balloons*, *Kendo*, *BookArrival*, *Newspaper*, *UndoDancer*, *GTFly*). Results were assessed against state of the art codecs with medium complexity, which excludes JPEG2000: JPEG and JPEGXR(Reference software 1.41) for lossy coding and JPEG-LS (version 0.6.4.1) for lossless coding. JPEG is a reference for image coding with low

complexity and it is widely used. JPEGXR has almost the same complexity as the LAR codec. It offers lossy and lossless coding but it has greatly reduced scalability. JPEG-LS is optimized only for lossless coding. It is important to note that the Best-Prediction technique based on the LAR codec framework achieves both lossy and lossless efficient coding.

5.1. Objective results

In a first set of experiments, objective tests were examined. We applied the proposed Best-Prediction technique on the LAR codec framework. By modifying the quantization parameter from 1 to 120 and setting the threshold to $Th = \frac{1}{3}Q$ initially and then $Th = \frac{2}{3}Q$, we generated the PSNR-bpp results of depth images shown in Fig. 6 and Fig. 7. The proposed technique outperforms the classic LAR by about 1-2 db depending on the bit-rate with 20%-30% bit saving. Subsequently, the resulting R-D performances were compared to the JPEG and JPEGXR encoders for lossy coding. The BP algorithm outperforms JPEG. At low bit-rates (lower than 0.1 bpp), the proposed joint depth/texture technique outperforms JPEGXR - up to 50% bit saving at very low bit-rates as shown in the zoomed images. For high bit-rates (PSNR higher than 45db), the proposed algorithm was outperformed by JPEGXR.

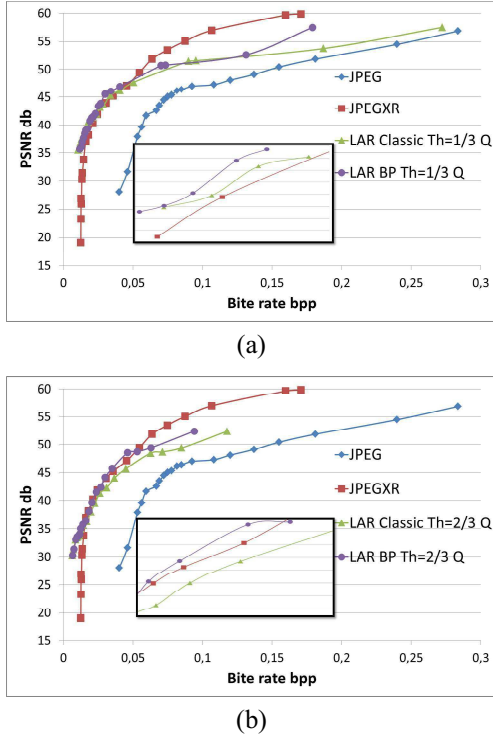


Fig. 6. Rate-Distortion curves of depth images for UndoDancer frame 250 view 1 with (a) $Th = \frac{1}{3}Q$, (b) $Th = \frac{2}{3}Q$

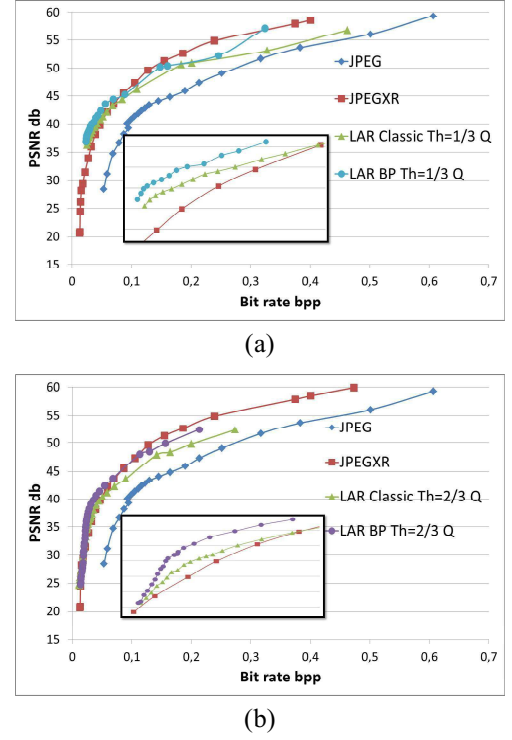


Fig. 7. Rate-Distortion curves of depth images for GTFLy frame 157 view 1 with (a) $Th = \frac{1}{3}Q$, (b) $Th = \frac{2}{3}Q$

For high quality synthesized videos in FTV and 3DTV, lossless coding can be preferred. Therefore, rate performances were compared to the lossless coding standard JPEGLS. The Best-Prediction scheme achieved a higher gain in comparison with classic LAR (36% gain for real images and 25% gain for virtual synthesized images) as shown in Table 1. Subsequently, JPEGLS slightly outperforms the proposed technique. It is important to note that the distortion type of the decoded depth image by LAR can be managed by the choice of homogeneity threshold. For $Th = \frac{2}{3}Q$, the Depth_Grid contains large blocks compared to $Th = \frac{1}{3}Q$, but involves a finer quantization.

Table 1. Rate in bpp of depth images coded in lossless mode with classic LAR and with proposed technique

Depth image	Rate (bpp)			
	LAR	LAR_BP	JPEGXR	JPEGLS
GTFLy	0.84	0.62	1.70	0.33
Balloons	1.10	0.73	1.85	0.44
Newspaper	1.21	0.89	1.97	0.52
BookArrival	1.45	0.92	2.36	0.50
UndoDancer	0.52	0.44	1.25	0.43

5.2. Visual results of decoded depth images

Even when the PSNR of depth images decoded by JPEGXR is higher than those decoded by LAR, we notice that the proposed technique enhances the visual quality of depth images in comparison with classic LAR and JPEGXR, (see Fig. 8).

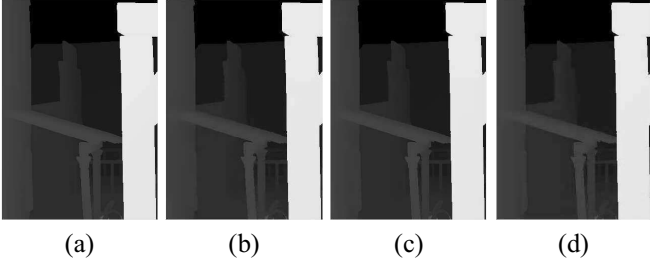


Fig. 8. Visual quality comparison for GTFLy view 1 frame 157 at 0.08 bpp over decoded (a) uncompressed depth image; (b) depth images coded by classic LAR (43 db, initial partition of 20227 blocks); (c) depth images coded by the proposed technique (44.65 db, initial partition of 25436 blocks); (d) depth images coded by JPEGXR (45.5 db).

5.3. Visual results of synthesized views

The final and more important issue in depth map coding is the visual quality of the resulting synthesized views. We used the View Synthesis Reference Software (VSRS 3.0) [18] to render intermediate views. In this set of experiments, we considered the original texture images and the decoded depth maps with $Th = \frac{2}{3}Q$, in order to evaluate the compression effect of the proposed technique on the synthesized views (Fig. 9).

It is clearly noticeable that the quality of intermediate views synthesized from depth images decoded by the proposed technique is much better than the one synthesized from depth images decoded by JPEGXR. The correlation between texture and depth information used to improve the prediction of the depth component, enables a higher quality for synthesized views (Fig.10, 11,12).

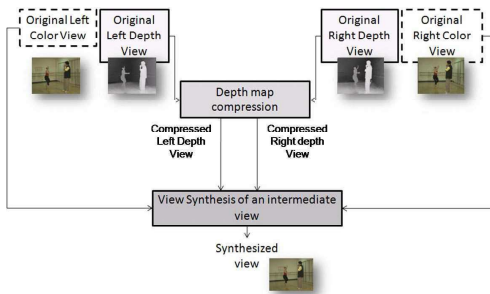


Fig. 9. View rendering framework

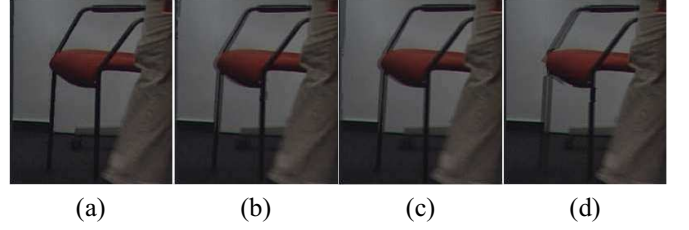


Fig. 10. Visual quality comparison of rendered view for BookArrival view 9 frame 033 at 0.012 bpp using (a) uncompressed depth image; (b) depth images decoded by classic LAR (26.6 db, initial partition of 1663 blocks); (c) depth images decoded by the proposed technique (25.7 db, initial partition of 1154 blocks); (d) depth images decoded using JPEGXR (21.5 db).

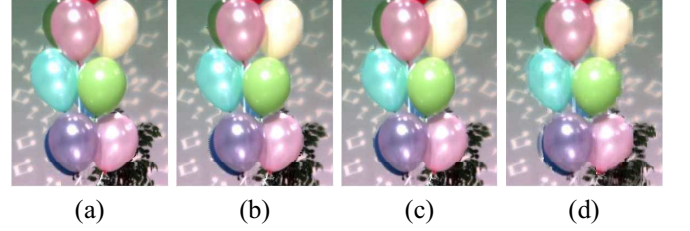


Fig. 11. Visual quality comparison of rendered view for Balloons view 4 frame 1 at 0.013 bpp using (a) uncompressed depth images; (b) depth images decoded by classic LAR (29.56 db, initial partition of 1894 blocks); (c) depth images decoded by the proposed technique (28.77 db, initial partition of 1450 blocks); (d) depth images decoded by JPEGXR (22.65 db).

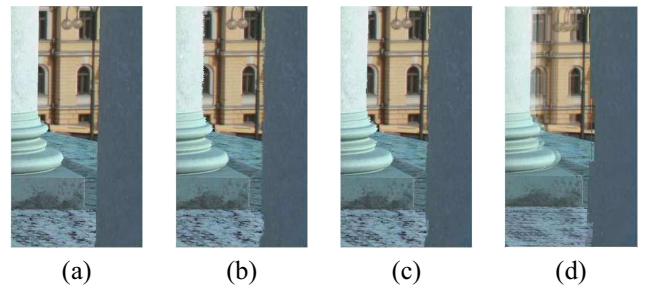


Fig. 12. Visual quality comparison of rendered view for UndoDancer view 3 frame 250 at 0.012 bpp using (a) uncompressed depth images; (b) depth images decoded by classic LAR (34.99 db, initial partition of 4867 blocks); (c) depth images decoded by the proposed technique (35.08 db, initial partition of 4867 blocks); (d) depth images decoded by JPEGXR (19.03 db).

6. CONCLUSION

In this paper, a novel coding scheme for depth data is proposed. The joint depth/texture algorithm presented here was built on an efficient and simple LAR coder framework, initially designed for 2D images. This technique does not code the depth as a component independent from the corresponding texture image, but exploits the correlation between texture and depth data to improve depth coding performances. The best predictor of the Luma component of the texture image, found a posteriori, is applied on the corresponding depth component. The proposed scheme was assessed against JPEG and JPEGXR for lossy coding and against JEGLS for lossless coding. According to objective results, the proposed scheme reduces the coded depth bit-rate in comparison with the classic LAR scheme. Although the Best-Prediction algorithm was outperformed by JEGLS for lossless coding, it achieves both lossy and lossless coding quite efficiently. Subjective results shows that the proposed scheme also improves the visual quality of synthesized intermediate views. Such scheme can be applied on multiple predictive coders. In future work we will focus on making a large set of subjective tests, and also we will work on finding an automatic algorithm that find the optimal pair of $\{Q; Th\}$.

7. ACKNOWLEDGEMENT

Authors thank AZM and SAADE Association for their fund of this research. This works is also supported by the European Celtic-Plus project H2B2VS and was partially funded by Finland, France, Spain, Switzerland and Turkey.

8. REFERENCES

- [1] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3d video and free viewpoint video - technologies, applications and mpeg standards," *IEEE ICME*, pp. 2161–2164, 2006.
- [2] C. Fehn and R. S. Pastoor, "Interactive 3-dtv-concepts and key technologies," *Proceedings of the IEEE*, vol. 94, pp. 524–538, Mar. 2006.
- [3] 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 CSVT*, vol. 17, pp. 1606–1621, Nov. 2007.
- [4] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," *IEEE ICIP*, vol. 1, pp. 201–204, 2007.
- [5] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Muller, P. With, and T. Wiegand, "The effect of depth compression on multiview rendering quality," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp. 245–248, 2008.
- [6] E. Bosc, M. Pressigout, and L. Morin, "Focus on visual rendering quality through content-based depth map coding," *PCS*, pp. 158–161, 2010.
- [7] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Miller, "Quality assessment of 3d video in rate allocation experiments," *IEEE ISCE*, pp. 1–4, 2008.
- [8] C. Lan, J. Xu, and F. Wu, "Improving depth compression in hevc by pre/post processing," *IEEE ICMEW*, pp. 611–616, 2012.
- [9] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient compression of multi-view depth data based on mvc," *3DTV Conference*, pp. 1–4, 2007.
- [10] S. Milani, P. Zanuttigh, M. Zamarin, and S. Forchhammer, "Efficient depth map compression exploiting segmented color data," *IEEE ICME*, pp. 1–6, 2011.
- [11] J. Lee, H. Wey, and D. Park, "A fast and efficient multi-view depth image coding method based on temporal and inter-view correlations of texture images," *IEEE Transactions on CSVT*, vol. 21, pp. 1859–1868, Dec. 2011.
- [12] F. Pasteau, C. Strauss, M. Babel, O. Deforges, and L. Bedat, "Adaptive colour decorrelation for predictive image codecs," *EUSIPCO*, Aug. 2011.
- [13] K. Samrouth, O. Deforges, and F. Pasteau, "Quality constraint and rate-distortion optimization for predictive image coders," *will be printed in Proc. - SPIE Conference on EI*, 2013.
- [14] 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," *TCSVT*, vol. 17, pp. 974–987, 2007.
- [15] M. Babel and O. Deforges, "Lossless and lossy minimal redundancy pyramidal decomposition for scalable image compression technique," *IEEE ICASSP*, vol. 3, pp. 249–252, 2003.
- [16] Y. Morvan, D. Farin, and P. With, "Depth-image compression based on an r-d optimized quadtree decomposition for the transmission of multiview images," *IEEE ICIP*, vol. 5, pp. 105–108, 2007.
- [17] S. Rane and G. Sapiro, "Evaluation of jpeg-ls, the new lossless and controlled-lossy still image compression standard, for compression of high-resolution elevation data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 2298–2306, Oct. 2001.
- [18] M. Tanimoto, T. Fuji, K. Suzuki, N. Fukushima, and Y. Mori, "Reference softwares for depth estimation and view synthesis," *ISO/IECJTC1/SC29/WG11, Archamps, France, Tech. Rep. M15377*, Apr. 2008.