Electronic Letters on Computer Vision and Image Analysis 15(1):14-26, 2016

Edge-aware wedgelet estimation for depth maps compression

Dorsaf Sebai, Faten Chaieb and Faouzi Ghorbel

Cristal Laboratoy, National School of Computer Science, University of Manouba, Tunisia

Received 18th Sep 2015; accepted 24th Mar 2016

Abstract

In recent years, Multi-view Video plus Depth (MVD) compression has received much attention thanks to its relevance to free viewpoint applications needs. An efficient compression, that causes the least distortion without excessive rate and complexity increase, becomes a must particularly for depth maps. These latter can be compressed efficiently by the 3D extension of High Efficiency Video Coding (3D-HEVC), which has explored wedgelets. Such functions lead to significant Rate-Distortion tradeoffs. However, they require a very large computational complexity involved by the exhaustive search used for the estimation of the wedgelet subdivision line. In this paper, we propose a rapid localization of this latter using an edge detection approach. The experimental results show that the proposed approach allows an important gain in terms of encoding delay, while providing competitive depth maps and synthesized views quality compared to the exhaustive search approach.

Key Words: depth maps, 3D-HEVC, wedgelets, subdivision line estimation, exhaustive search, edge detection, synthesized views.

1 Introduction

Three-dimensional video (3D video) has been subject to promising and increasing interests in many innovative applications such as 3D TeleVision (3DTV) and Free Viewpoint Video (FVV). First, 3DTV allows the viewer to perceive the depth of the scene thanks to multiview stereoscopic and autostereoscopic display systems. Second, the FTV provides the ability for users to interactively navigate and select a viewpoint in the video scene. Despite their efficiency for the 3DTV domain, the Classical Stereoscopic Video (CSV) and MultiView Video (MVV) have rapidly presented limits for the FTV. The lack of geometry information prohibits view synthesis and rendering of intermediate views necessary to FTV applications. To deal with this drawback, recent researches have concentrated on depth maps that allow a 2D representation of a 3D scene. They associate, to each texture pixel, a depth value that represents the distance of this latter to the capture camera. Based on depth information, virtual views can be reconstructed employing Depth-Image-Based Rendering (DIBR) methods. In this context, Video plus Depth (V+D) and Multiview Video plus Depth (MVD) present main video formats that exploit the depth maps favor. However, view synthesis from V+D produces large disocclusion regions in rendered

Correspondence to: <sebaidorsaf@yahoo.fr>
Recommended for acceptance by <Joan Serra-Sagrist>
http://dx.doi.org/10.5565/rev/elcvia.807
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

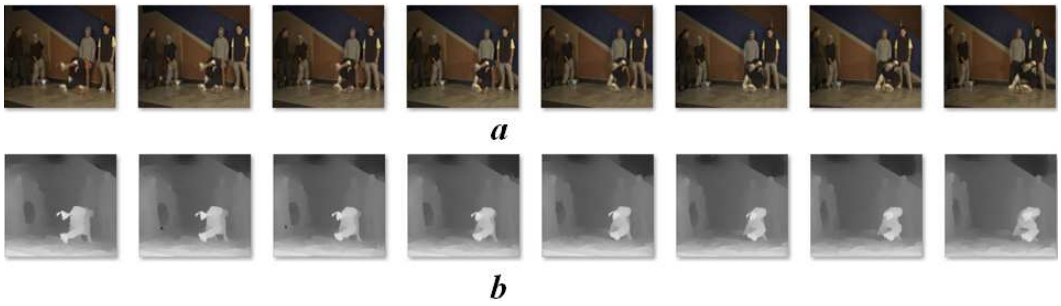

Figure 1: 8 views captured by 8 different cameras : a) Texture images and b) Corresponding depth maps.

views. Thus, the navigation is limited to viewpoints that are very close to the captured one. This shortcoming is particularly reduced with MVD since the same scene is captured by several cameras, allowing a wider range of viewpoints. For each of the above-mentioned 3D video formats, a coding standard has been addressed : H264/MPEG-4 AVC (Advanced Video Coding) [15] for CSV, H264/MPEG-4 MVC (Multiview Video Coding) [13] for MVV and MPEG-C Part 3 [14] for V+D. 3D-High Efficiency Video Coding (3D-HEVC), subject of this paper, is the ongoing standard destinated to MVD compression [11].

2 Depth maps compression : State of the art

As shown in figure 1, Multiview Video plus Depth (MVD) [12] includes sequences of texture images and their corresponding depth maps. These latter are bi-dimensional gray level images representing the distance of each texture pixel to capture camera. Recent efforts point toward an efficient coding that preserves depth maps particularities, namely their piece-wise planar conception and the critical impact of pixels near contours on perceptual quality of synthesized views. First, a depth map can be approached to a piece-wise planar signal where adjacent plans are separated by arbitrary shaped contours. Every plan corresponds to an object of the scene while contours, placed at object boundaries, reproduce sharp discontinuities between foreground and background objects. Second, coding errors of depth discontinuities produce visible degradation of synthesized views [1]. However, it has been shown that in low texture regions of a scene, errors in depth have limited effect on view synthesis quality.

Since depth images are used for view synthesis and are not themselves displayed, some efforts aim at reducing depth maps coding artifacts that cause severe distortion of synthesized views. Cheung et al. [17] define ”Don’t Care Regions” (DCRs), for each pixel, where a depth value outside the DCR will lead to a synthesis distortion larger than a threshold value. Then, they perform sparsification of the depth map in an orthogonal basis, optimally trading off its representation sparsity and its adverse effect on synthesized view distortion. More recently, this idea is reused by Cheung et al. [16] replacing DCRs by penalty function. For each pixel, a quadratic penalty function is defined based on sensitivity of interpolated images to pixel depth values during rendering process. Transform domains used in [17] [16] are classical orthogonal basis that represent dictionaries of minimum size, concentrating the signal energy over a set of few vectors. However, vectors sets larger than basis, particularly redundant dictionaries, are needed to build sparse representations of complex signals.

Instead of predicting the depth compression effect on synthesis quality, many coding research work aim at faithfully reconstruct depth maps. Maitre et al. [18] propose a codec that relies on a lifting implementation of Shape-Adaptive Discrete Wavelet Transform (SA-DWT). SA-DWT independently treats surfaces separated by edges which, and unlike classical wavelet transforms, provides much sparser decomposition with small coefficients along depth discontinuities. Furthermore, Shen et al. [19] present a new set of Edge-Adaptive Transform (EAT) as an alternative to the classical Discrete Cosine Transform (DCT). EAT avoids filtering across depth discontinuities and so avoids creating large coefficients. However, transform domains used in [18] [19] need an encoded representation of major edge locations to be shared between both encoder and decoder

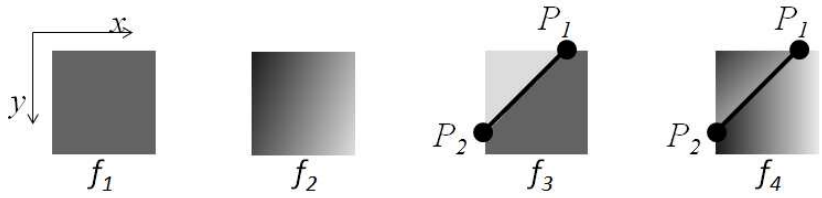

Figure 2: Depth maps compression modeling functions : f_1) *Constant function* $f_1(x, y) = \alpha$, f_2) *Linear function* $f_2(x, y) = \alpha + \beta x + \gamma y$, f_3) *Wedgelet function* 2 *constant functions* $+ P_1$ *and* P_2 *coordinates* and f_4) *Platelet function* 2 *linear functions* $+ P_1$ *and* P_2 *coordinates.*

sides. Morvan et al [3] exploit the linear piece-wise nature of platelet and wedgelet functions, to detail in next section, in order to approximate depth planar surfaces separated by shaped edges.

3 Wedgelet-based depth maps compression

In this section, we pay attention to Morvan et al [3] work since the wedgelet representation is retained for 3D-HEVC standard. Morvan et al [3] exploit the planar piece-wise nature of platelet [10] and wedgelet [9] functions to approximate depth maps. Morvan et al. perform a recursive segmentation of the depth map into a quadtree. Obtained depth blocks are then estimated using four modeling functions f_1, f_2, f_3 and f_4 (see figure 2). The constant function f_1 represents a depth block with a single coefficient (α) that simply corresponds to the average of the depth values of the block pixels. f_2 is a linear function of three coefficients (α, β, γ) adequate for depth blocks that contain a gradient. Wedgelet (f_3) and platelet (f_4) functions capture depth discontinuities by aligning a subdivision line along object boundaries, such that the two resulting sub-regions can be approximated by constant or linear functions. Besides modeling coefficients of wedgelet and platelet functions, coordinates of the subdivision line extremities are also determined.

The wedgelet concept proposed in [3] has been recently implemented into the 3D-HEVC-based proposal [11]. In fact, this latter includes, besides the conventional HEVC modes, four different Depth Modeling Modes (DMM) [4] based on wedgelets. They differ in their way to partition the depth blocks. The ”Explicit wedgelet signalization” consists in finding and transmitting the best approximation of a depth block through a wedgelet partition. The wedgelet block partition information is stored in the form of a binary partition pattern, signaling which pixel belongs to each of the two wedgelet sub-regions. This implies an extensive search of the best wedgelet partition using the original depth of the current block to be coded. In fact, the patterns of all possible combinations of line partitions are stored in a Look-Up Table (LUT). An exhaustive search is then carried out within the LUT to select the best approximating partition for the block. In the ”Intra-predicted wedgelet partitioning”, the separation line of the current block is predicted from its neighborhood by continuing the separation line in the current block from a neighboring wedgelet reference block. The ”Inter-component-predicted wedgelet partitioning” aims to explore the redundancy between the two MVD components, namely texture and depth. It typically consists in predicting the wedgelet partition of the current depth block from a co-located texture one in the MVD. The wedgelet partition is not transmitted but signaled so that the inter-component prediction uses the reconstructed texture video as reference for the partitioning. The last mode, ”Inter-component-predicted contour partitioning”, is similar to the third one. The difference consists in predicting a contour partition instead of a line one.

In this paper, we propose a new depth modeling approach based on edge detection in order to save compression delay of the ”Explicit wedgelet signalization” mode of 3D-HEVC. Typically, we aim to reduce the depth maps encoding time with a better, at worst sustained, 3D video quality than the original mode. Quality is measured on depth maps as well as on synthesized views.

The rest of the paper is organized as follows. In Section 4, we present the proposed approach. Section 5

presents experimental results and analysis. Section 6 concludes the work.

4 Proposed method

We first highlight the high computational cost of the wedgelet exhaustive search approach. We then propose an edge-based one.

4.1 Exhaustive search limit

To estimate the subdivision line of wedgelet function in "Explicit wedgelet signalization" mode of 3D-HEVC for a given depth block, a LUT (L) is constructed. L includes all possible subdivision lines that can divide the depth block into two sub-blocks. The cardinality of L ($\#(L)$) for a given depth block with height h and width w is depicted in equation (1).

$$\#(L) = w^2 + h^2 + 4wh - 7w - 8h + 9 \tag{1}$$

An exhaustive search is next performed through the large set L in order to select the best fitting subdivision line. This latter is selected such that it minimizes the approximation error between the model and the original depth block. One major disadvantage of this greedy search is its computational complexity that inhibits its implementation on low-cost embedded systems and even on regular platforms. This consequently highly affects the depth maps encoding delay that is one of the most important criteria required for multiview video coding methods [2]. Instead of exhaustive searching within all possible separating lines, we propose a new approach based on edge detection. We particularly propose an edge detector usage for subdivision lines estimation. This narrows the subdivision line search to only depth discontinuities, and an exhaustive search through all possible wedges is no more required. We also propose to restrict wedgelet modeling partition to single-edge blocks, i.e. blocks that include one edge.

4.2 Edge-based approach

As shown in figure 3, the proposed approach proceeds in three steps: Sobel edge detection, Freeman chain construction and orientation test. Edges at object boundaries are first detected. In the context of depth maps, the localization of sharp depth discontinuities should be quasi-perfect since they affect synthesis quality. For that, we make use Sobel detector that, at the opposite of other gradient operators, does not rely on noise prefiltering before local maxima edge detection. This inhibits edge smoothing and makes it less vulnerable to contour localization errors. This accurate contour localization is however at the cost of a noisy edge over-detection. To cope with this, contours that are very short are excluded. In fact, short contours might be useless to the reconstruction and might unnecessarily increase the edge coding cost. Once the Sobel operator is performed, the Freeman chain of detected edges is constructed using compressed polygonal segments, such that only extremities of edges are leaved. The delineating subdivision line corresponds to the straight line that connects edge extremities. We finally perform an orientation test in order to determine whether a pixel belongs to the left or right sub-region of a depth block.

Based on the number of edges per Freeman chain, only single-edge depth blocks are estimated by wedgelet function. We particularly restrict depth blocks modeling using wedgelets to only depth blocks that contain a single edge. In fact, it is insignificant to use piece-wise planar functions to estimate smooth and multi-edges depth blocks. The former does not contain object boundaries which limits its adequate modeling to a constant value, i.e. the mean value of the block pixels depths. The latter gathers multiple depth discontinuities that require further subdivisions of the block to smaller Coding Units (CUs) until more homogeneous areas are reached. In order to explicit the idea, two depth blocks examples are illustrated in figure 4. Sobel edge detector distinguishes one edge in the depth block of figure 4a. This single-edge block is easily and efficiently approximated by wedgelet function. Its subdivision line corresponds to the straight line that connects pixels

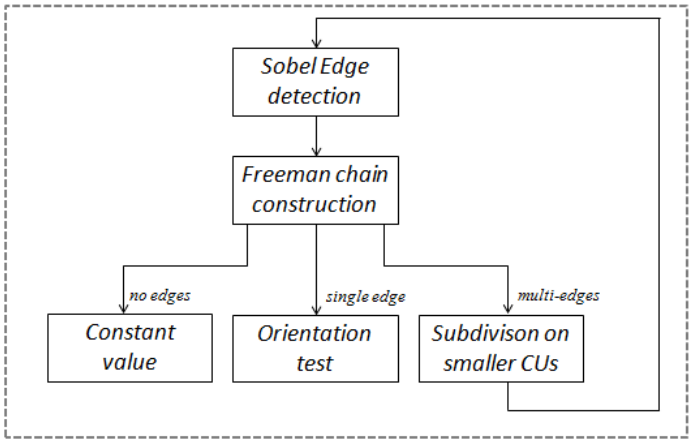

Figure 3: Flowchart of the proposed approach.

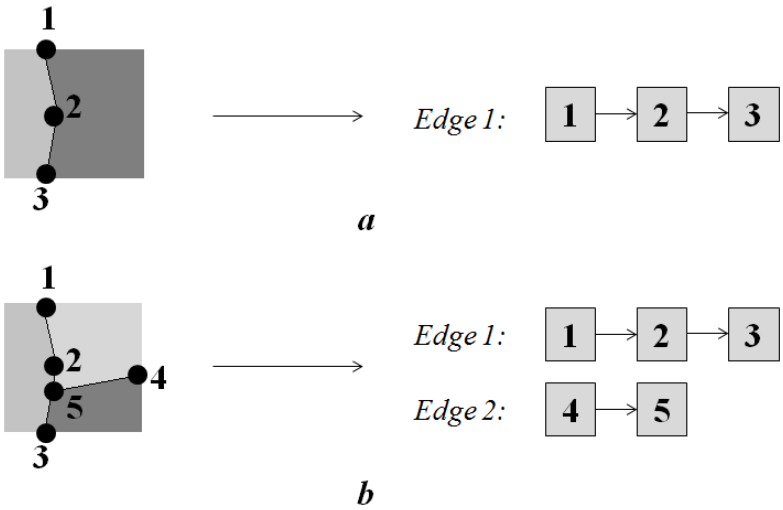

Figure 4: Depth blocks and their corresponding Freeman chains: a) Single-edge depth block and b) Multi-edges depth block.

1 and 3. At the opposite, the Freeman chain of the block of figure 4b includes 2 linked lists, one per each edge. This hampers the modeling of this multi-edges block by a single subdivision line. A further subdivision is solicited to have a better estimation of the block.

5 Results and analysis

All experiments are performed using the 3D-HEVC Test Model (3D-HTM version 4.1) [5]. Comparisons concern the original "Explicit wedgelet signalization" of 3D-HEVC using the Wedgelet Exhaustive Search ($W + ES$) and the proposed Edge Detection based approach ($W + ED$). Concerning test sequences, we make use of three benchmark depth sequences, namely $Breakdancers$ and $Ballet$ of Microsoft [6] and $Champagne$ of Nagoya University [20] (see figure 5). $Breakdancers$ sequence presents depth maps with weak discontinuities. $Ballet$ and $Champagne$ sequences include on the contrary depth maps with stronger edges, i.e. greater difference of intensity between the pixels at each side of the edge.

Figure 5: First frames of first views of the test depth sequences : $Breakdancers$ (left), $Ballet$ (middle) and $Champagne$ (right).

5.1 Experiment 1 : Compression delay

Table 1 compares the average per frame compression delay spent by $W + ES$ and $W + ED$. As can be observed, $W + ED$ implemented as described in section 4 is noticeably faster than $W + ES$, i.e. nearly 7 times faster in average. The poor runtime scaling for $W + ES$ is due to the exhaustive search of the wedgelet subdivision lines, e.g. a 16 $\times$ 16 depth block requires a LUT of more than 1300 possible subdivisions (see equation (1)). Another interesting feature, not discussed in this paper, is that the proposed approach do improve the processing requirements in terms of memory and processor.

Table 1: Average per frame compression delay (in seconds) of test sequences encoded by $W+ES$ and $W+ED$ approaches.

	W+ES	**W+ED**
Breakdancers	83	12
Ballet	95	16
Champagne	119	18
Mean	99	15

5.2 Experiment 2 : Depth maps quality evaluation

The aim of this experiment is to compare the performances of $W + ES$ and $W + ED$ approaches in terms of depth maps quality. Since PSNR is a pure mathematical metric, we propose to use a new full-reference measure, SSIMplus [8]. It provides real-time prediction of the perceptual quality of a video based on Human Visual System (HVS) behaviors, video content characteristics, e.g. spatial and temporal complexity and video resolution, display device properties, e.g. screen size, resolution, and brightness, and viewing conditions, e.g. viewing distance and angle. Compared to most popular and widely used quality assessment measures, SSIMplus has shown a higher perceptual quality prediction accuracy and closer performances to Mean Opinion Scores [8].

Table 2 summarizes SSIMplus values measured for encoded depth maps at two bitrates, namely 0.01 bpp and 0.1 bpp. The evaluation is performed at two different bitrate values that correspond to two critical cases, namely low and high bitrates.

Table 2: SSIMplus values of test depth maps encoded using $W + ES$ and $W + ED$ approaches at 0.01 *bpp* and 0.1 *bpp*.

	0.01 bpp		*0.1 bpp*	
	W+ES	**W+ED**	**W+ES**	**W+ED**
Breakdancers	41	43	47	48
Ballet	43	44	50	50
Champagne	44	46	49	51
Mean	43	44	49	50

Besides SSIMplus HVS-based measure, figures 6 and 7 allow visual evaluation of areas zoomed from encoded depth maps of test sequences at 0.01 *bpp* and 0.1 *bpp*. It can be noticed that the proposed method can reduce some distortions along object boundaries than $W + ES$, particularly at low bitrate. It allows a more refined discontinuities approximation and details preservation instead of the coarse angular modeling generated by $W + ES$. In addition to the Sobel edge detector precision, this can be explained by the restriction of wedgelet estimation to single-edge depth blocks. As already explained in Section 4.2, a multi-edges block is subdivided in our case until homogeneous blocks, i.e. single-edge blocks, are obtained. This is not the case for The "Explicit wedgelet signalization" mode of 3D-HEVC that inherits the same quadtree coding structure [21] for both texture and depth components, since the texture and its associated depth represent the same scene at the same time instant and viewpoint. Then, the depth quadtree is limited to the coded texture quadtree. This can affect the modeling of a multi-edges depth block, particularly if its corresponding texture block can be efficiently predicted with a large coding structure, e.g. a Coding Tree Unit (CTU) [21] of 32×32 pixels.

5.3 Experiment 3 : Synthesized views quality evaluation

Since the main use of depth maps is in view synthesis operations, these experimentations are concerned with the evaluation of views that can be synthesized from already compressed depth images. Typically, the following experimentations consist in coding left and right views from $Breakdancers$, $Ballet$ and $Champagne$ sequences. The decoded views are then used for view synthesis using View Synthesis Reference Software (VSRS) [7] of Nagoya University. We notice that original texture images are used for view synthesis because our goal is to evaluate distortions caused by depth coding.

Table 3 shows SSIMplus results of candidate methods obtained for synthesized images of test sequences encoded at 0.01 *bpp* and 0.1 *bpp*. We also present zoomed synthesized regions (*see.* figures 8 and 9). Compared to $W + ES$ mode, the proposed method achieves competitive visual synthesis quality. It even allows the reconstruction of some details with less harmful distortions, particularly at 0.01 *bpp*.

Table 3: SSIMplus values of test synthesized views obtained from original textures and depth maps encoded using $W + ES$ and $W + ED$ approaches at 0.01 *bpp* and 0.1 *bpp*.

	0.01 bpp		*0.1 bpp*	
	W+ES	**W+ED**	**W+ES**	**W+ED**
Breakdancers	29	30	38	38
Ballet	32	34	43	44
Champagne	39	41	47	48
Mean	33	35	42	43

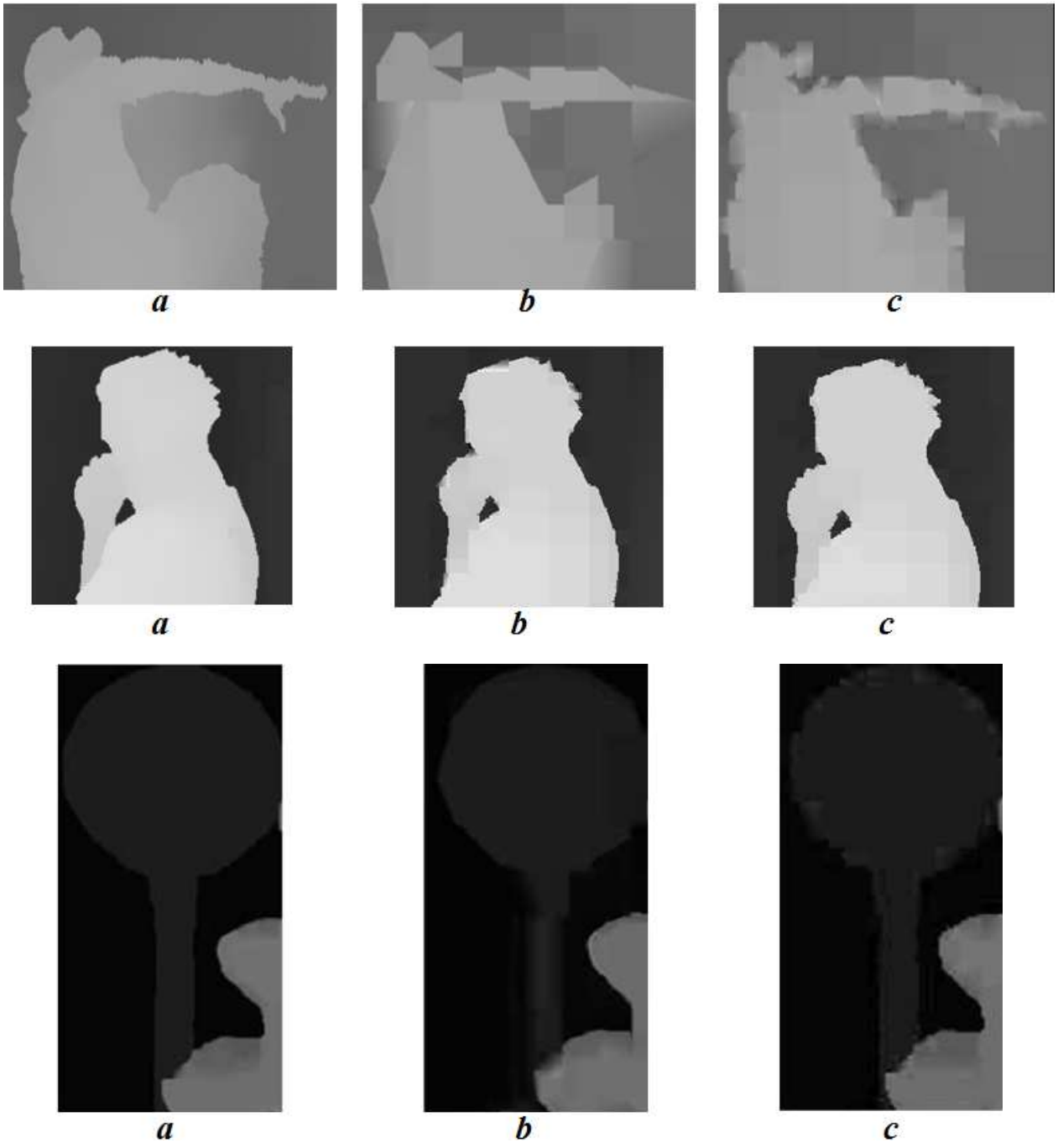

Figure 6: Zoomed areas of test depth maps : (a) original, encoded at 0.01 bpp using (b) $W + ES$ and (c) $W + ED$.

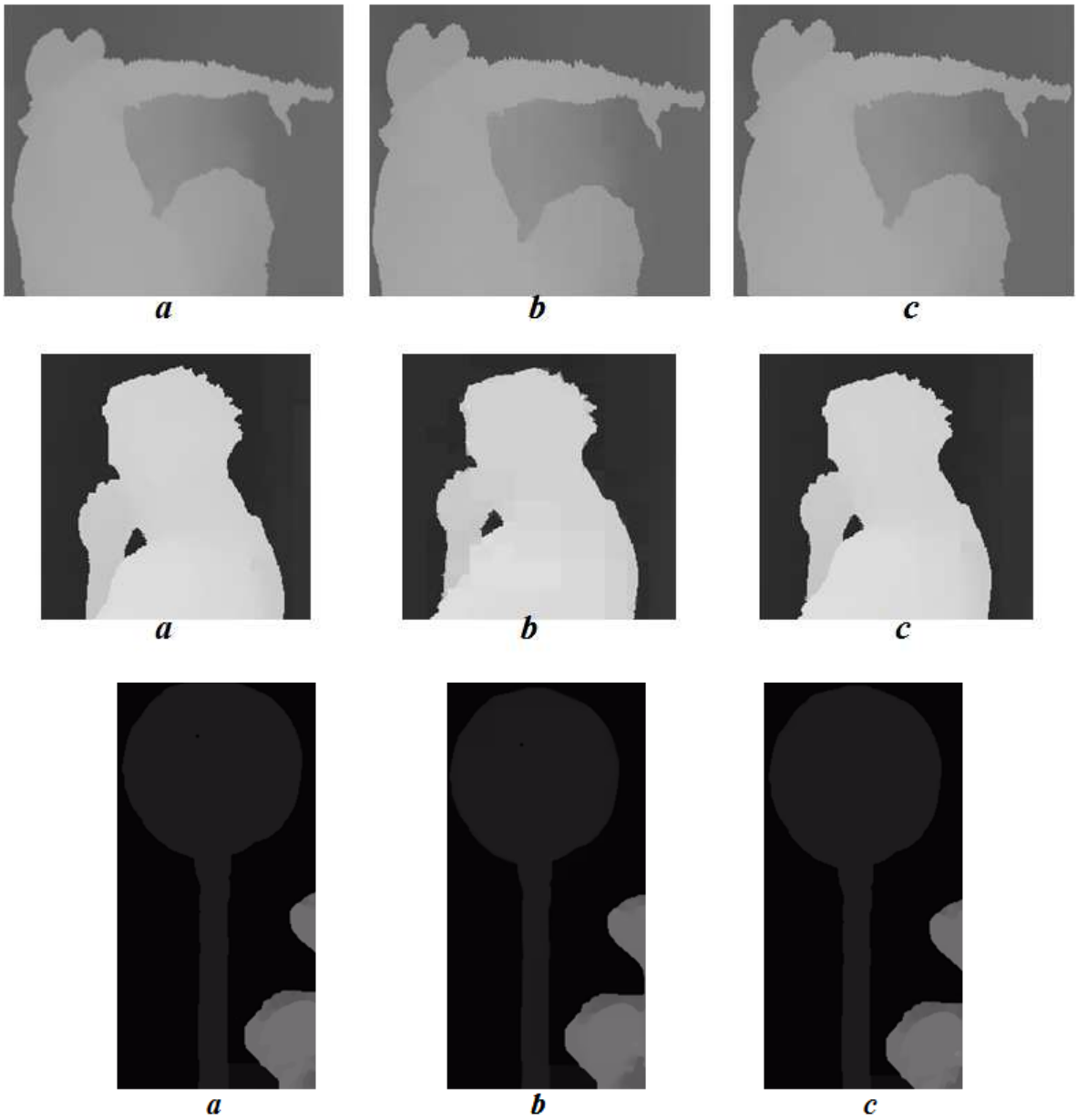

Figure 7: Zoomed areas of test depth maps : (a) original, encoded at 0.1 *bpp* using (b) $W+ES$ and (c) $W+ED$.

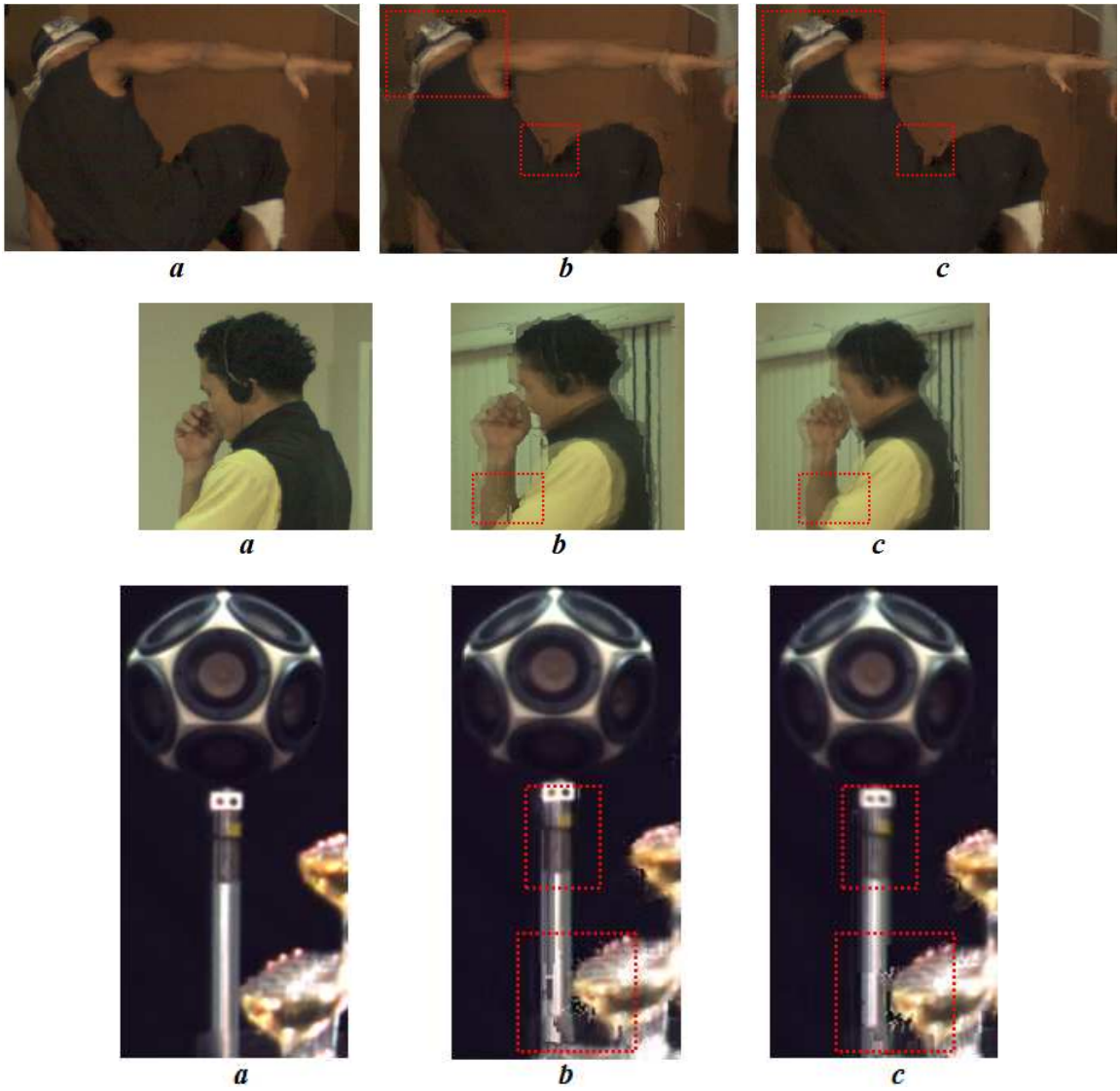

Figure 8: Zoomed areas of test sequences : (a) original, synthesized from original textures and depth maps encoded at 0.01 *bpp* using (b) $W + ES$ and (c) $W + ED$.

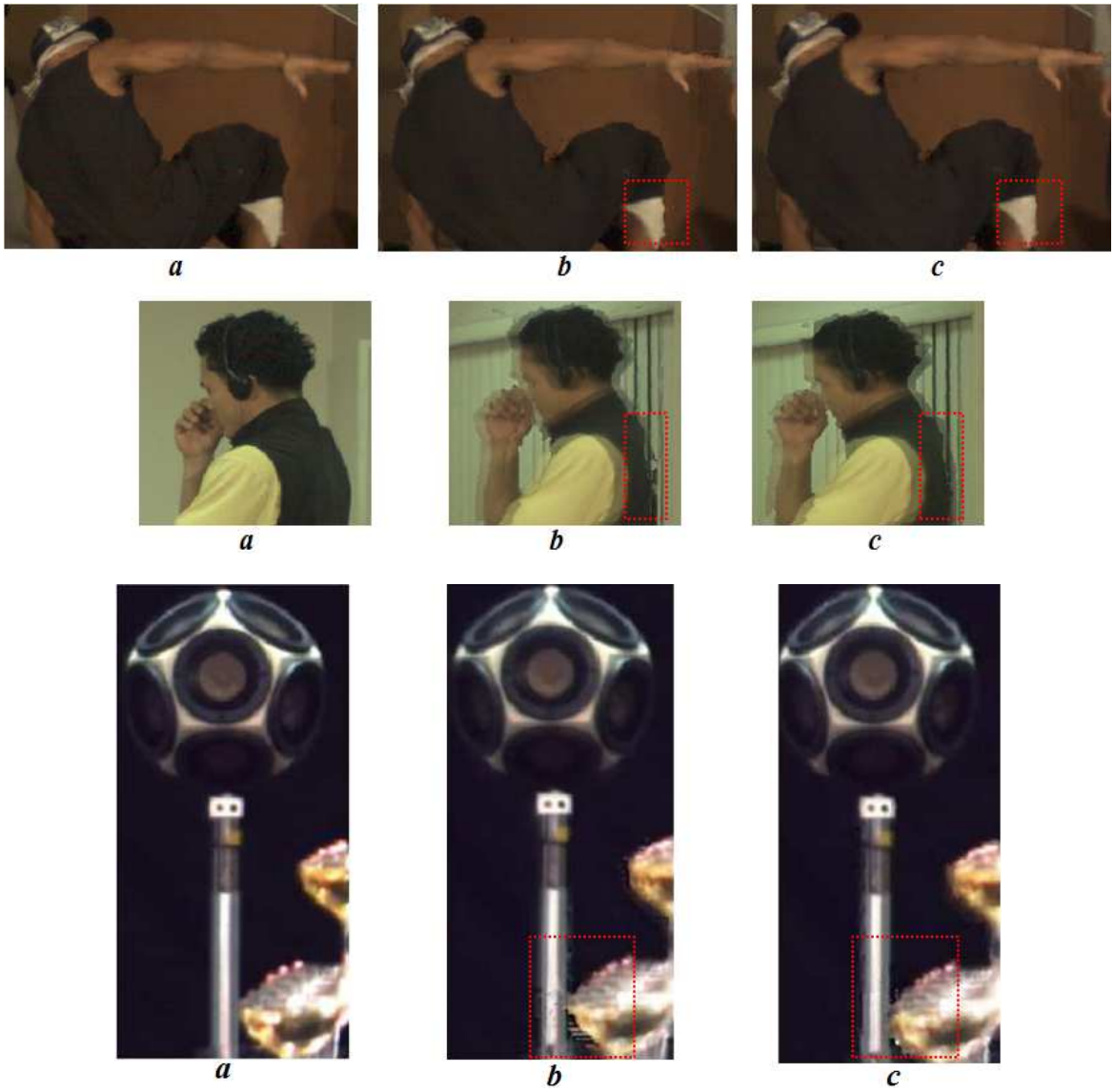

Figure 9: Zoomed areas of test sequences : (a) original, synthesized from original textures and depth maps encoded at 0.1 *bpp* using (b) $W + ES$ and (c) $W + ED$.

6 Conclusion

We implemented an edge-aware searching of wedgelet subdivision line for depth maps compression in 3D-HEVC. Our aim was to improve the searching speed of the original wedgelet partitions exhaustive search. The experimental results show that the proposed method can reduce the compression delay more than 6 times, while preventing depth discontinuities quality degradation. The HVS-based SSIMplus metric and visual evaluation show that the proposed method can reach competitive object boundaries preservation than the original modeling mode in depth maps as well as in synthesized views.

References

[1] W.-S. Kim, A. Ortega, P. Lai, D. Tian, C. Gomila, "Depth map distortion analysis for view rendering and depth coding," *International Conference on Image Processing*, 2009. DOI:10.1109/ICIP.2009.5414304

[2] ISO/IEC JTC1/SC29/WG11: "Requirements on Multiview Video Coding v.7", 2006.

[3] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Muller, P.H.N.de With, T. Wiegand, "The effects of multi-view depth video compression on multiview rendering", *Signal Processing: Image Communication Journal* 24(1-2):7388, 2008. DOI:10.1016/j.image.2008.10.010

[4] P. Merkle, C. Bartnik, K. Muller, "3D video : Depth coding based on inter-component prediction of block partitions," *Picture Coding Symposium*, 2012. DOI:10.1109/PCS.2012.6213308

[5] https ://hevc.hhi.fraunhofer.de/svn/svn 3DVCSoftware/tags/HTM-4.1.

[6] http ://research.microsoft.com/en-us/downloads/5e4675af-03f4-4b16-b3bca85c5bafb21d/.

[7] M. Tanimoto, T. Fujii an K. Suzuki, N. Fukushima, Y. Mori, "Reference softwares for depth estimation and view synthesis", *ISO/IEC JTC1/SC29/WG11 MPEG2008/M15377*, France, 2008.

[8] Z. W. A. Rehman, K. Zeng, "Display device-adapted video quality-of-experience assessment", *Electronic Imaging, Human Vision and Electronic Imaging XX*, 2015. DOI:10.1117/12.2077917

[9] D. Donoho, "Wedgelets : nearly minimax estimation of edges", *Annals of Statistics* 27(3):859897, 1999. DOI:10.1214/aos/1018031261

[10] R. M. Willett, R. D. Nowak, "Platelets : a multiscale approach for recovering edges and surfaces in photon-limited medical imaging", *IEEE Transactions on Medical Imaging* 22(3):332350, 2003. DOI:10.1109/TMI.2003.809622

[11] K. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, H. F. Rhee, G. Tech, M. Winken, T. Wiegand, "3D high efficiency video coding for multi-view video and depth data", *IEEE Transactions on Image Processing* 22(9):3366-3378, 2013. DOI:10.1109/TIP.2013.2264820

[12] P. Merkle, A. Smolic, K. Muller, T. Wiegand, "Multi-view video plus depth representation and coding," *International Conference on Image Processing*, 2007. DOI:10.1109/ICIP.2007.4378926

[13] P. Merkle, A. Smolic, K. Muller, T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Transactions on Circuits and Systems for Video Technology* 17(11):14611473, 2007. DOI:10.1109/TCSVT.2007.903665

[14] A. Bourge, J. Gobert, F. Bruls, "MPEG-C PART 3 : Enabling the introduction of video plus depth contents," *IEEE Workshop on Content Generation and Coding for 3D Television*, 2006.

[15] T. Wiegand, G.J. Sullivan, G. Bjontegaard, A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology* 13(7):560576, 2003. DOI:10.1109/TCSVT.2003.815165

[16] G. Cheung, J. Ishida, A. Kubota, A. Ortega, "Transform domain sparsification of depth maps using iterative quadratic programming," *International Conference on Image Processing*, 2011. DOI:10.1109/ICIP.2011.6115673

[17] G. Cheung, A. Kubota, A. Ortega, "Sparse representation of depth maps for efficient transform coding," *Picture Coding Symposium*, 2010. DOI:10.1109/PCS.2010.5702491

[18] M. Maitre, M. N. Do, "Depth and depth-color coding using shape-adaptive wavelets," *Journal of Visual Communication and Image Representation* 21(5-6):513522, 2010. DOI:10.1016/j.jvcir.2010.03.005

[19] G. Shen, W.-S. Kim, S. K. Naran, A. Ortega, L. Jaejoon, W. HoCheon, "Edge-adaptive transforms for efficient depth map coding," *Picture Coding Symposium*, 2010. DOI:10.1109/PCS.2010.5702565

[20] http://www.tanimoto.nuee.nagoya-u.ac.jp.

[21] P. Helle, S. Oudin, B. Bross, D. Marpe, M. Bici, K. Ugur, J. Jung, G. Clare, T. Wiegand, "Block merging for quadtree-based partitioning in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology* 22(12):1720-1731, 2012. DOI: 10.1117/12.945932