Light-field video coding using geometry-based disparity compensation

Published in
3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), July 2-4,2014, Budabest, Hungary

DOI
10.1109/3DTV.2014.6874724

LIGHT-FIELD VIDEO CODING USING GEOMETRY-BASED DISPARITY COMPENSATION

Caroline Conti[1,2], *Péter Tamás Kovács*[3,4], *Tibor Balogh*[3], *Paulo Nunes*[1,2], *Luís Ducla Soares*[1,2]

[1]Instituto de Telecomunicações, [2]University Institute of Lisbon (ISCTE-IUL), Lisbon, Portugal
[3]Holografika, Budapest, Hungary, [4]Department of Signal Processing, Tampere University of Technology, Tampere, Finland

ABSTRACT

One of the main challenges in 3D light-field imaging approaches lies in the massive amount of visual information involved in providing 3D content with sufficient resolution. Consequently, adequate coding tools are essential for efficient transmission and storage of this type of content. In this context, this paper presents and evaluates two coding solutions based on the High Efficiency Video Coding (HEVC) scheme and for efficient compression of the 3D light-field content. These two coding schemes aim to exploit the 3D geometry-based disparity information in the 3D light field content and replace the block-based disparity estimation. In the first scheme, the disparity map of each view is used to directly derive the vectors for compensation, and in the second scheme these disparity vectors (for all views) are calculated (for non-occluded areas) from the disparity map of the base view. A comparative study of these proposed coding schemes is performed and future research directions are also discussed.

***Index Terms* —** Light-field, HoloVizio, HEVC, 3D video coding

1. INTRODUCTION

One of the main challenges in 3D light-field imaging approaches lies in the massive amount of visual information to provide 3D content with sufficient resolution. For instance, as opposed to transmission of 2 views, as seen in current 3DTV systems, Holografika's HoloVizio displays [*1*] require 100+ views as input. Consequently, adequate coding tools are essential for efficient transmission and storage of this large amount of data involved.

In this context, this paper presents and evaluates two coding solutions based on the HEVC scheme for efficient compression of 3D light field content. These coding schemes aim to exploit the 3D geometry-based disparity information in the 3D light field content to replace the disparity estimation through block-based matching. In the first proposed scheme, referred to as *MV direct-DV*, the disparity map of each view is used to directly derive the vectors for disparity compensation. In the second scheme, referred to as *MV base-DV*, the disparity information is only coded and transmitted for the Intra coded base view. Then, for the Inter coded views, the disparity information is derived from the disparity map of the base view by using multiview camera geometry. A comparative analysis of both schemes is performed to better understand the rate-distortion tradeoffs between disparity vector calculation and motion estimation.

The remainder of the paper is organized as follows: Section 2 provides an overview of HoloVizio technology and its 3D light-field content representation; Section 3 presents the two proposed coding schemes based on HEVC; Section 4 performs the evaluation of proposed coding schemes; and, finally, Section 5 concludes the paper and present some future work directions.

2. LIGHT-FIELD DISPLAYS AND 3D CONTENT REPRESENTATION

HoloVizio light-field (LF) 3D displays are able to present 3D imagery over a wide Field Of View (FOV) - currently up to 180º, with high resolution in all directions. As such, they require significant amount of image information from the represented scene, which needs to be captured/generated over a wide viewing angle that matches the represented viewing angle, often in the range of hundreds of views. Although LF displays do not output the 3D light field with a view-based approach, there is a mapping from views to LF slices that are then composed by the display's optics to form a continuous 3D image with motion parallax. Therefore, for the purposes of exploring video coding approaches, it is possible to assume that a coding method capable of generating the sufficient number of views in the end is sufficient.

Although storing and transmitting hundreds of views or LF slices directly is possible, more efficient methods based on depth/disparity maps and subsequent views synthesis on the receiver's side have been proposed and are under development and standardization. In fact, motion vectors used in hybrid video coding approaches are somewhat related to disparity when considering a multiview sequence of adjacent cameras (or a camera pan, for that matter), however the objective of motion estimation algorithms is to find the best matching blocks, which do not necessarily correspond to the real object movement caused by camera displacement. The motion estimation process might find very similar blocks which have no relation to real motion in the scene whatsoever, which then cannot be reused in subsequent views, assuming linear camera motion. However, if the motion vectors of adjacent views are coherent with the real scene depth, it is possible to assume that motion in subsequent frames will correspond to the actual disparity between adjacent views.

Although estimating depth/disparity for live scenes is a challenging task, it is easy to get ground truth depth data for synthetic scenes. During the rendering process of synthetic 3D content, the precise depth map is generated as side information to handle occlusions in the scene. Assuming lambertian surfaces, it is expected that depth values can be used as a basis for motion vectors.

3. PROPOSED CODING SCHEMES FOR LIGHT-FIELD CONTENT CODING

Since disparity/depth maps comprise geometric information of the 3D scene from a particular viewpoint, it is possible to use this information to exploit the redundancy between different views of the scene. For example, the current Multiview Video Coding (MVC) standard [*2*] employs a block-based disparity

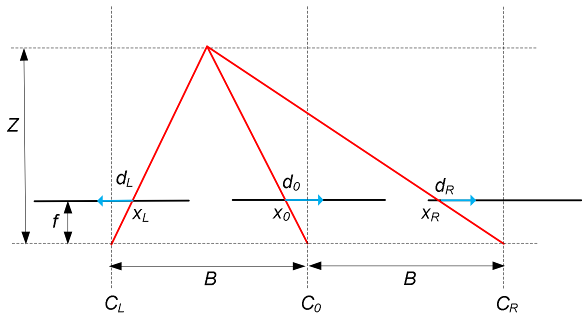


Figure 1 Disparity vector derivation considering a parallel camera geometry

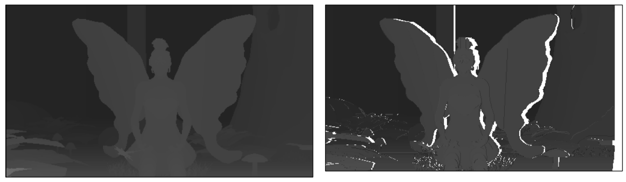
estimation technique to remove this existing inter-view redundancy. Thus, a disparity vector is derived and represents the displacement of the objects between the different views. However, if it is possible to correctly derive the disparity vectors from the real 3D geometry, the adjacent view image can be accurately estimated, and the block-based disparity estimation process can be skipped. Therefore, the goal of both proposed schemes is to analyze and better understand if the direct disparity vector derivation is comparable – in terms of rate-distortion gains – with the conventional disparity estimation process.

In the proposed *MV direct-DV* coding scheme, the motion estimation process is replaced by a direct 3D geometry-based disparity vector calculation. In this approach, a disparity vector with 8-bit precision is derived from the depth map of each 4×4 texture block. Considering only the first access unit (notice that temporal prediction was not considered in this analysis), the Base View is encoded using the conventional HEVC Intra prediction modes. Hence, when Inter coding a block of an Enhancement View, all the prediction block (PB) partition patterns existing in Inter prediction modes of HEVC [***3***] are enabled, but they are modified to integrate the direct disparity vector calculation. Finally, the encoder selects the best prediction mode in a Rate Distortion Optimization (RDO) sense.

However, it should be noticed that the *MV direct-DV* scheme does not take into account the existence of occluded regions where there is no reference block, e.g., the reference block corresponding to the calculated disparity vector falls outside the reference view picture.

The proposed *MV base-DV* coding scheme intends to handle the occluded regions issue. In this case, the disparity information is only coded and transmitted for the Intra coded view (Base View). Similar to the *MV direct-DV* scheme, each disparity vector is taken for each 4x4 texture block. For this, a modified-Intra frame is defined, where each PB is coded by using the existing Intra prediction modes of the HEVC, but the disparity vectors are now included in the prediction information. The Advanced Motion Vector Prediction (AMVP) scheme of HEVC [***3***] is enabled in this modified-Intra frame to efficiently encode these disparity vectors. Then, the disparity vector difference and the index to the list of predictor vector candidates are also entropy-coded and transmitted as in conventional HEVC.

Before encoding an enhancement view, the disparity information is derived from the disparity map of the base view by using multiview camera geometry. Some coding schemes in the literature have used this derivation process to efficiently predict the disparity vectors in a MVC coding approach and speedup the disparity and the motion estimation [***4***] [***5***]. In these approaches, the base disparity information is through conventional block-based disparity estimation. It should be noticed, however, that in the particular case of the MV base-DV scheme, the idea is to efficiently compress the disparity information to enable encoding and transmission of the disparity information that cannot be directly derived (corresponding to occluded areas).



(a) (b)

Figure 2 Example of a derived disparity map: (a) disparity map of the base view; (b) derived disparity map for a right side view, highlighting the occluded areas in white

In the *MV base-DV* coding scheme, a parallel camera geometry is considered, where each pair of adjacent cameras are equally spaced. This ensures that the vertical component of the disparity vector is always zero and the displacement between two matching points in different view images is simply defined by the difference between these points. Moreover, an “IPPP..” prediction structure is also considered, where each Inter coded enhancement view uses the (adjacent) previously coded view as reference.

An example of the considered camera geometry model is shown in Figure 1, where C_o, C_R and C_L represent three equally spaced cameras and the base view corresponds to the view captured by C_o.

In the condition of no occlusion, given a disparity vector d_o in a position x_o of the base view’s disparity map, it is possible, for instance, to derive the disparity vector d_R in the corresponding matching point, x_R, of the right view. In this case, the corresponding matching position, x_R, is displaced in relation to x_0 by the relative position, $\alpha_{0,R}$, between base view and of the current view (in this case, it corresponds to the relative position between C_o and C_R) multiplied by the absolute disparity vector amplitude $|d_o|$, i.e.,

$$x_R = x_0 + \alpha_{0,R} \times |d_o| \quad (1)$$

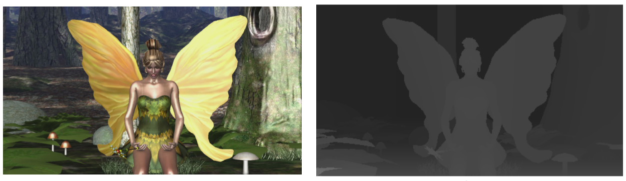
In this case, $\alpha_{0,R}$, corresponds to a negative scale of -1 and then, the matching point x_R is shifted to the left by $|d_o|$. Additionally, the relative disparity value, d_R, is given by the absolute disparity value, d_0, scaled by the relative position, $\beta_{R,0}$, between the current view and the used reference view in the “IPPP..” coding structure (in this case, the reference view is also C_o), i.e.,

$$d_R = \beta_{R,0} \times |d_0| \quad (2)$$

And, in this case, $\beta_{R,0} = -\alpha_{0,R}$.

Similarly, given d_o at x_o, it is also possible to derive the disparity vector, d_L, for the matching point, x_L x_L , in the left view. Thus, it will end up with $\alpha_{0,L} = -\alpha_{0,R}$ and $\beta_{L,0} = -\beta_{R,0}$ since right and left views are in opposite directions. Generally, a significant number of disparity vectors for a given view can be simply derived from the disparity map of another view. However, there are two cases where the derived disparity vector could not be considered: *i)* when the derived disparity vector corresponds to a reference block position that falls outside the reference view; *ii)* when there is no valid disparity since the corresponding position is in an occluded area. An example of such derived disparity map is shown in Figure 2b, where the white values illustrate the existing occluded areas.

Therefore, when encoding the current enhancement view, an early Skip mode is used for blocks where a valid disparity vector is derived. This means that the conventional skip mode of HEVC is changed to only consider one predictor vector candidate (i.e., the one derived from the base view disparity map). Thus, if there is a valid disparity vector candidate, this skip mode is automatically selected. On the other hand, for the missing blocks (i.e., where the disparity vector can’t be considered) all prediction modes existing in an Inter coded frame of HEVC are allowed and the best prediction mode is selected in a RDO sense.



(a) (b)

Figure 3 Test sequence *Fairy Tale*: (a) texture of central view; (b) corresponding disparity map relative to the left adjacent view.

4. EXPERIMENTAL RESULTS

This section assesses the rate-distortion (RD) performance of the proposed coding schemes and discusses some relevant results. For this, a multiview sequence called *Fairy Tale* with 61 views with a resolution of 1920×1080 pixels is used as well as their corresponding disparity maps. The central view texture and corresponding disparity map of this test sequence can be seen in Figure 3. The proposed coding schemes were implemented using the HEVC reference software HM9.1. In order to evaluate the *MV direct-DV* scheme, three different scenarios were tested and compared:

a1) *HEVC MVD All Intra*: This scenario represents the simulcast case for independently encoding multiview texture plus disparity map (referred to as MVD) using HEVC. In this case, each disparity map has a quarter of the resolution of the texture. Each view and the corresponding disparity map is coded with the HEVC reference software HM 9.1 using the "Intra, main" configuration defined in [***6***];

b1) *HEVC MVD*: This scenario represents a MVD coding scheme based on HEVC. Disparity maps are independently encoded. Each texture view is coded with the HEVC reference software HM 9.1 using the "Low-delay P, main" configuration (defined in [***6***]) considering an "IPPP.." with Group of Pictures equal to 1 and Intra Period equal to 31. In the first intra period the central view is Intra coded (first I frame), then all left views are Inter coded (P frames) with central to left order. In the second intra period, the coding order becomes from central to right;

c1) *MV direct-DV*: This scenario represents the proposed *MV direct-DV* coding scheme based on HEVC, where the disparity estimation is replaced by the direct 3D geometry-based disparity vector calculation. The "Low-delay P, main" configuration [***6***] with the same "IPPP.." prediction structure as in b1) is used. In this case, two ways to calculate the disparity vectors for each 4×4 where considered: *i)* by using the mean of all 16 disparity vectors (referred to as Mean); and *ii)* by using the most common vector in the 16 disparity vectors (referred to as Mode).

The RD performance for the *MV direct-DV* scheme is presented in Figure 4, where the PSNR Y values correspond to the luminance Peak Signal-to-Noise Ratio (PSNR) for the encoded texture views and the bits correspond to the overall bits spent for coding both texture and depth data. The eight RD points correspond to quantization parameter (QP) values 22, 26, 27, 31, 32, 36, 37 and 41. From these results, it can be seen that the *MV direct-DV (Mode)* presents a better RD performance compared with the *MV direct-DV (Mean)*, therefore only the most common vector calculation was considered in the *MV base-DV* scheme. Figure 5 shows the residual information for the first right P coded view using the *HEVC MVD* and the *MV direct-DV (Mode)* schemes. From Figure 4 and Figure 5, it can be concluded that the significant lower performance of the *MV direct-DV*, when compared with the *HEVC MVD* is mainly due to the higher residual information in occluded regions when coded with *MV direct-DV* scheme.

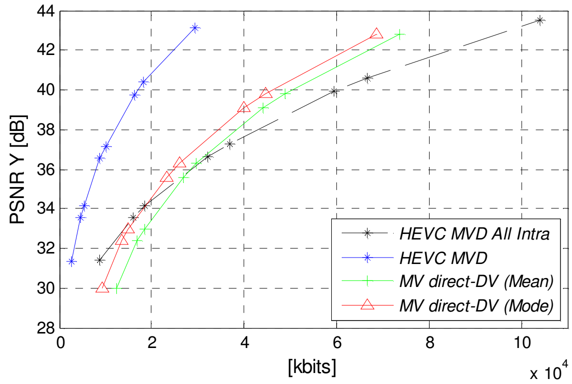
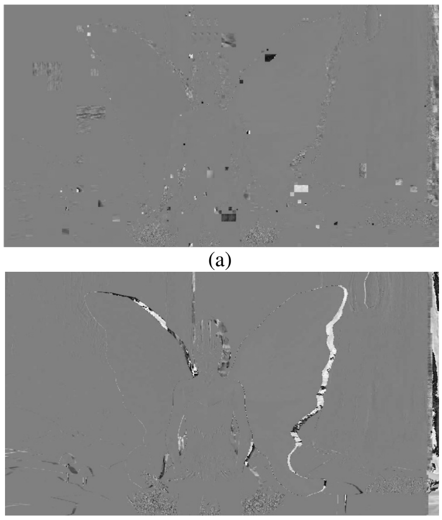


Figure 4 RD performance for the proposed *MV direct-DV* coding scheme



(a)

(b)

Figure 5 Residual luminance for the first inter coded right view using: (a) *HEVC MVD*; and (b) *MV direct-DV (Mode)*

To evaluate the *MV base-DV* coding scheme, three different scenarios were tested and compared:

a2) *HEVC MVC All Intra*: This scenario represents the simulcast case for a multiview video coding based on HEVC. In this case, each view is independently encoded with the HEVC reference software HM 9.1 using the "Intra, main" configuration [***6***];

b2) *HEVC MVC*: This scenario represents a MVC coding scheme based on HEVC. Each view is coded with the HEVC reference software HM 9.1 using the "Low-delay P, main" configuration [***6***] considering an "IPPP.." with Group of Pictures equal to 1 and Intra Period equal to 31. The view coding order used to evaluate the *MV direct-DV* coding scheme is also adopted in this case;

c2) *MV base-DV*: This scenario represents the proposed *MV base-DV* coding scheme based on HEVC. The "Low-delay P, main" configuration with the same "IPPP.." prediction structure as in b2) is used. The disparity vectors, which are sent in the Intra coded views, are calculated for each 4×4 block (through the mode of 16 disparity vectors).

Notice that the disparity information is not considered for the first two scenarios (a2) and (b2), since in the *MV base-DV* it is only sent for the central view. Therefore, these results should be interpreted as preliminary results, which will guide the proposal of future improved coding schemes. It should be also noticed that, in this case, the disparity information that cannot be derived from the base view shall also be coded and transmitted to be available at the decoder side (such that further views can be better interpolated).

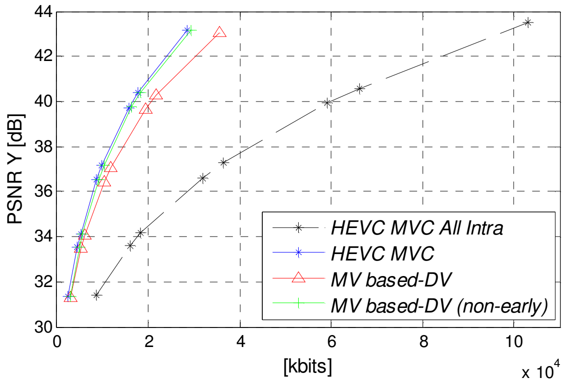


Figure 6 RD performance for the *MV base-DV* coding scheme.

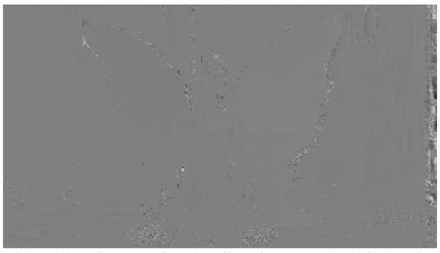


Figure 7 Residual luminance for the first inter coded right view using the *MV base-DV* coding scheme

The RD performance for the *MV base-DV* coding scheme is presented in Figure 6. The eight RD points correspond to quantization parameter (QP) values 22, 26, 27, 31, 32, 36, 37 and 41. It can be seen that the *MV base-DV* is significantly more advantageous than the *MV direct-DV*. However it is still outperformed by the *HEVC MVC* scheme. Moreover, Figure 6 shows that this happens since the *MV base-DV* spent more bits than *HEVC MVC*. Nevertheless, analyzing the residual information shown in Figure 7, it can be seen that the *MV base-DV* scheme presents significant less residual information even when compared with the *HEVC MVC* (Figure 6In order to understand the reason for the lower performance of the *MV base-DV* (notably, the increased number of bits spent in this case), a further analysis was employed and is presented in Figure 8. The objective of this analysis was to compare the performance of the proposed modified-Intra coding with the independent Intra coding of texture and disparity vectors. Hence, all the views were Intra coded using *MV base-DV* (referred to as *MV base-DV All Intra*) and compared with the *HEVC MVD All Intra* scenario. In Figure 8, the presented results consider the average luminance PSNR of the encoded texture views and the overall bits spent for coding the both texture and depth. From these results, it can be concluded that the lower RD performance of the *MV base-DV All Intra*, when compared to *HEVC MVD All Intra*, indicates that it is necessary to improve the performance of the modified-Intra coded views so as to obtain a better overall performance for the *MV base-DV* (Figure 6). Possible solutions to improve the Intra coding performance may include improving the vector prediction scheme, or not including the disparity information in the Intra coded views (and encoding it separately).

Alternatively, Figure 6 also shows the performance of the *MV base-DV* scheme when the early skip mode is disabled and the encoder can choose (in a RD optimization sense) whether to use the derived disparity vector or to the estimate the "true" disparity. This scenario is referred to as *MV base-DV (non-early)* in Figure 6. As can be seen in this figure, the performance of the *MV base-DV (non-early)* is comparable with *HEVC MVC*. Therefore, it is expected that by combining the *MV base-DV (non-early)* with a better Intra coding scheme, it would be possible to improve overall performance and outperform *HEVC MVC*.

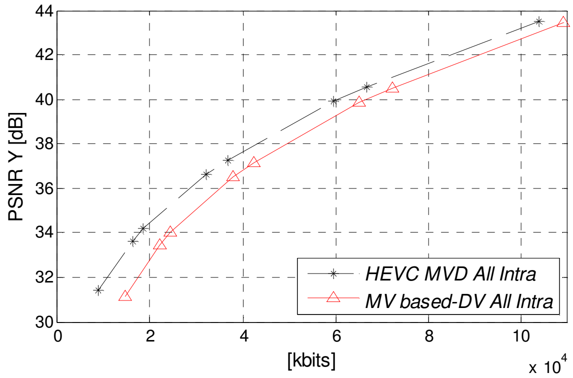


Figure 8 RD performance for the *MV base-DV* coding scheme considering an "All Intra" configuration

5. FINAL REMARKS

This paper proposes to study and evaluate coding solutions based on the High Efficient Video Coding for efficient compression of 3D light-field content. For this, two alternative coding schemes were proposed and evaluated that aimed to directly exploit the 3D geometry-based disparity information in 3D light field content. The proposed *MV base-DV* coding scheme was shown to be more advantageous, as it was able to considerably reduce the residual information in occluded areas. However, further improvements in this scheme are still needed to improve the performance of Intra-coded frames.

6. ACKNOWLEDGEMENTS

This work was supported by 3D-CounTourNet Cost Action (IC1105) under the Short Term Scientific Mission COST-STSM-IC1105-13972.

The research leading to these results has received funding from the PROLIGHT-IAPP Marie Curie Action of the People programme of the European Union's Seventh Framework Programme, REA grant agreement 32449.

7. REFERENCES

[1] Zoltán Megyesi, Attila Barsi, and Tibor Balogh, "3D Video Visualization on the HoloVizio System," in *3DTV Conference 2008*, Istanbul, Turkey, 2008, pp. 269-272.

[2] A. Vetro, T. Wiegand, and G.J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626-642, April 2011.

[3] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649-1668, December 2012.

[4] A. Avci et al., "Efficient disparity vector coding for multi-view 3D displays," in *Three-Dimensional Image Processing (3DIP) and Applications*, San Jose, 2010, pp. 1-7.

[5] B.W. Micallef, C.J. Debono, and R.A. Farrugia, "Exploiting depth information for fast motion and disparity estimation in Multi-view Video Coding," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON 2011)*, May 2011, pp. 1-4.

[6] Frank Bossen, "Common HM test conditions and software reference configurations," Doc. JCTVCL1100 2013.