

This material is published in the open archive of Mid Sweden University

DIVA <http://miun.diva-portal.org>

to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

Schwarz, S.; Olsson, R.; Sjöström, M.; Tourancheau, S. , "Adaptive depth filtering for HEVC 3D video coding," *Picture Coding Symposium (PCS), 2012*, pp.49-52, 07-09 May 2012

<http://dx.doi.org/10.1109/PCS.2012.6213283>

© 2012 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Adaptive depth filtering for HEVC 3D video coding

Sebastian Schwarz, Roger Olsson, Mårten Sjöström, and Sylvain Tourancheau
Department of Information Technology and Media, Mid Sweden University
Holmgatan 10, 85170, Sundsvall, Sweden
Phone: + (46) 060 1488-62, Fax: + (46) 060 1488-30,
Email: [sebastian.schwarz, roger.olsson, marten.sjostrom, sylvain.tourancheau]@miun.se

Abstract—Consumer interest in 3D television (3DTV) is growing steadily, but current available 3D displays still need additional eye-wear and suffer from the limitation of a single stereo view pair. So it can be assumed that auto-stereoscopic multiview displays are the next step in 3D-at-home entertainment, since these displays can utilize the Multiview Video plus Depth (MVD) format to synthesize numerous viewing angles from only a small set of given input views. This motivates efficient MVD compression as an important keystone for commercial success of 3DTV. In this paper we concentrate on the compression of depth information in an MVD scenario. There have been several publications suggesting depth down- and upsampling to increase coding efficiency. We follow this path, using our recently introduced Edge Weighted Optimization Concept (EWOC) for depth upscaling. EWOC uses edge information from the video frame in the upscaling process and allows the use of sparse, non-uniformly distributed depth values. We exploit this fact to expand the depth down-/upsampling idea with an adaptive low-pass filter, reducing high energy parts in the original depth map prior to subsampling and compression. Objective results show the viability of our approach for depth map compression with up-to-date High-Efficiency Video Coding (HEVC). For the same Y-PSNR in synthesized views we achieve up to 18.5% bit rate decrease compared to full-scale depth and around 10% compared to competing depth down-/upsampling solutions. These results were confirmed by a subjective quality assessment, showing a statistical significant preference for 87.5% of the test cases.

I. INTRODUCTION

Auto-stereoscopic displays for three-dimensional television (3DTV) render intermediate views from the Multiview Video plus Depth (MVD [1]) format to reduce the required transmission bandwidth. Upcoming video compression, i.e. High Efficiency Video Coding (HEVC [2]), is assumed to be the future standard for this transmission. The question is how we can combine the characteristics of MVD with the strengths of HEVC to further decrease bandwidth?

Stereoscopic 3D content is making its way from the cinemas into our homes. But the raised consumer expectations in a living room scenario call for a more sophisticated 3DTV representation without additional eye-wear and increased viewing-angle. This need can be satisfied with auto-stereoscopic multiview displays, which provide spatial-multiplexed views for a range of viewing positions. Transmitting each single view is inappropriate, since intermediate views can be generated from an MVD input. Still, the required bandwidth is much higher than in traditional (2D) TV broadcast. Efficient MVD coding is the key to realize 3DTV broadcasting and guarantee its commercial success.

This paper focuses upon coding MVD depth map sequences. Depth maps describe scene geometry with per-pixel depth values represented as a gray-scale image. The European AT-TEST project [3] has shown that depth maps can be efficiently compressed by state-of-the-art video coding, allocating 10-20% of the overall bit rate budget for depth map sequences [4]. Klimaszewski et al. show that coding efficiency can be increased by transmitting downsampled depth maps [5]. Several approaches have been proposed to enhance depth upscaling results with corresponding texture information, e.g. utilizing texture edges [6] or weighting color similarity [7], [8]. Kopf et al. introduced Joint Bilateral Upscaling (JBU) [9], a concept widely used in sensor-fusion upscaling, e.g. for time-of-flight (ToF) range sensors, which can also be transferred into a depth map coding scenario.

Recently we introduced an Edge Weighted Optimization Concept (EWOC) for ToF super-resolution [10], utilizing edge information from a video feed. Just like JBU this approach can easily be transferred into a depth map coding scheme. The novelty in this paper is that we do not only use texture information for upscaling, but also identify unimportant depth regions before downsampling. By exploiting the characteristics of depth maps we apply pre-processing to these regions, improving HEVC coding efficiency with only fractional loss in synthesis quality.

The remainder of this paper is organized as follows: We propose a novel depth map coding scheme Sec. II, describe its evaluation in Sec. III and present the results in Sec. IV. The paper is concluded in Sec. V.

II. PROPOSED METHOD

The scope of this paper lies in depth compression for MVD transmission, where HEVC will play an important role as the new upcoming video coding standard. We utilize the efficiency of HEVC and concentrate on depth pre- and post-processing to increase coding efficiency. The scheme proposed in this paper produces fully HEVC-compliant depth maps for a single view. Current MVD coding actions, like inter- and intra-view prediction, are not part of this paper but fully supported, since we apply no changes to the video file structure. Fig. 1 shows the principle of our depth coding proposal. Starting with uncompressed texture image V and depth map D , we apply an adaptive low-pass filter on the depth map. The resulting depth map D_{adapt} is downsampled and encoded with HEVC. After decompression we apply EWOC depth upscaling using

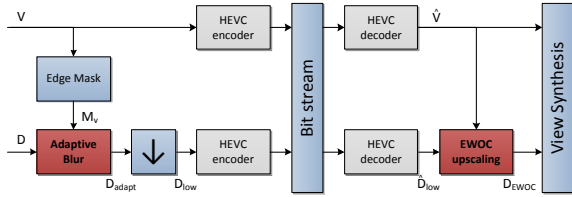


Fig. 1. Proposed depth coding scheme.

information from the decoded full resolution texture image \hat{V} to gain the full resolution depth map D_{EWOC} for view synthesis. The novelty of this approach is distributed over the two blocks marked in red: Adaptive Blur and EWOC upscaling. This section gives a detailed description of these processing steps as well as the applied downsampling scheme.

A. Adaptive depth blurring

In this step we make use of the characteristics of depth maps and combine them with the strengths of modern video coding. Depth maps consist of large smooth regions and sharp value transitions between regions. Video coding algorithms are very good in compressing uniform areas by removing redundancy. If we increase the uniformity in smooth depth regions, while still keeping important information at region transitions. We can increase the coding efficiency with a fractional loss in quality. We assume that important depth transitions between objects are represented by edges in the texture image and generate an edge map from the video frame V using a 'Canny' edge detector. For the later downsampling step we widen the edges in the edge map to form the edge mask M_V . This is a binary mask describing 'important' object transitions with ones and uniform regions with zeros. Using M_V we convolute depth map D with the Gaussian lowpass filter G_{16} to remove potential high frequencies in constant depth regions, while keeping the important information at object borders. The non-uniformly blurred depth map D_{adapt} is gained as follows:

$$D_{adapt} = (D * G_{16}) \cdot ((1 - M_V) * G_3) + D \cdot (M_V * G_3), \quad (1)$$

where G_{16} and G_3 are both Gaussian lowpass filters. G_{16} has a size of 16x16 pixel and $\sigma = 4$. G_3 is 3x3 pixel with $\sigma = 0.5$. The edge mask M_V is convoluted with G_3 to smooth the transitions between filtered and non-filtered parts of the depth map.

B. Depth downsampling

Spatial subsampling is frequently used in image and video compression schemes and is a popular trend for depth compression. Due to the large, only gradually changing areas in depth maps, it provides good bit rate savings with only limited loss in quality. The problem with traditional downsampling is the significant edge blurring due to lowpass anti-aliasing filters. This inflicts the exact depth values needed at object boundaries. We reduce this edge-blurring effect by applying an anti-aliasing filter with a cut-off frequency above the Nyquist

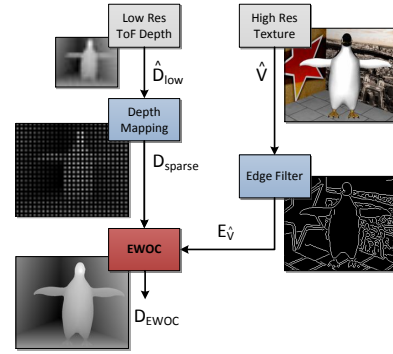


Fig. 2. Edge Weighted Optimization Concept (EWOC) for depth map upscaling.

frequency. In this way frequencies just below the Nyquist frequency are preserved, accepting some aliasing.

$$D_{low}(x, y) = D_{adapt}(x \cdot n, y \cdot n) * G_3 \quad (2)$$

For the evaluated sequences, G_3 provides sufficient anti-aliasing before downsampling by a factor of $n = 2$.

C. EWOC depth upscaling

After transmission, we need to reconstruct the decoded subsampled depth map \hat{D}_{low} to the original full resolution for pixel-by-pixel view synthesis. We apply EWOC depth upscaling, utilizing edge information from the corresponding texture frame. The basic principle of the EWOC is shown in Fig. 2: Values of the low resolution depth map \hat{D}_{low} are rearranged on a full resolution grid according to their exact spatial position, resulting in the sparse depth map D_{sparse} . We fill D_{sparse} solving a least square error problem, where we encourage the depth of each pixel $d(x, y)$ to be similar to its spatial neighbors. To avoid depth blurring at object boundaries, we introduce a weighting function Q_E , allowing pixels on object borders to be less similar. For Q_E we take the assumption from Sec. II-A that objects borders are represented by the edge map $E_{\hat{V}}$ from the decoded video frame \hat{V} :

$$Q_E(x, y) = 1 - E_{\hat{V}}(x, y) \quad (3)$$

The spatial smoothness requirements for the edge weighted optimization can then be described as follows:

$$Q_E(x, y) \cdot (d(x, y) - d(x + 1, y)) = 0 \quad (4)$$

$$Q_E(x, y) \cdot (d(x, y) - d(x, y + 1)) = 0 \quad (5)$$

A more detailed explanation of the whole upscaling process can be found in the paper introducing EWOC [10].

III. TEST ARRANGEMENT AND EVALUATION CRITERIA

This new coding scheme has been evaluated on the test sequences 'Street' and 'Hall2' from Poznań University of Technology, two photographic 1920x1088 pixel resolution sequences with estimated depth maps [11]. This choice was motivated by the quality of their estimated depth maps, their open availability for the research community and the different

depth structure in both sequences. While 'Hall2' has large uniform areas with straight edges, 'Street' is more erratic with small details and a more complex depth structure. Both sequences are also part of the evaluation criteria for the recent MPEG call for proposals (CfP) on 3D Video Coding Technology [12]. This allows for an easy comparison to upcoming trends in 3D video compression.

A. Test procedures

To show the validity of our approach, two testing procedures have been designed: The first to assess the feasibility of adaptive blurring in a depth coding scheme, the second to point out its benefits in MVD coding.

For the first test view synthesis with EWOC upscaled depth maps was compared to synthesis using full resolution and JBU upscaled depth maps, respectively. For EWOC the depth maps were pre-processed as described in Sec. II-A, both EWOC and JBU depth sequences were downsampled as described in Sec. II-B. No pre-processing was applied to the JBU sequences. We chose an up-/downscaling factor of $n=2$, which leads to a 4:1 reduction in the number of depth values. The low resolution depth maps were then upscaled using JBU and EWOC in two different tests. The depth map sequences were encoded with the quantization parameters (QP) 16, 20, 24, 28, 32, 40 and 44. The video sequences with a fixed QP 32.

In the second test we utilized the achieved depth bit rate reduction to increase the bit rate budget of the texture sequences for an increased video quality. We used the bit rate anchor points for a 2-view HEVC compliant coding scenario from the MPEG CfP [12] resulting in QPs as shown in Tab. I. Note that any additional overhead, e.g. camera and depth parameters, have been excluded from the bit rate allocation. The overall maximum bit rate was divided by 2 for left and right view and distributed between texture and depth. For the coded full-scale depth maps we allocated 20% of the bit rate budget to the depth sequences, for our approach only 10%, giving the same total bit rate.

The results of the second test scenario have been additionally evaluated by a subjective quality assessment following ITU recommendations [13]. We asked 20 test subjects, aged between 19 and 49, to rate the quality of synthesized views at the eight compression anchors presented in Tab. I. We presented four training sequences followed by the eight test sequences on a 23" full HD resolution display. The background luminance was set at 15% of the display luminance and the viewing distance was at three times the display height. The subjects were asked to rate in a pairwise stimulus-comparison (SC) scheme at the matching overall bitrates. The perceived quality difference between our approach and full-scale HEVC was rated on a scale from 'much better' (+3), 'better' (+2), 'slightly better' (+1), 'similar' (0), 'slightly worse' (-1), 'worse' (-2) to 'much worse' (-3).

B. Methodology

Video and depth sequences were coded with the latest HEVC test model HM-4.0 [14]. We used a GOP size of 12

TABLE I
HEVC QPs FOR MPEG CfP OVERALL BIT RATE ANCHORS [12].

Sequence	point 1	point 2	point 3	point 4
Hall2 V_{full}	QP 48	QP 44	QP 41	QP 36
Hall2 D_{full}	QP 50	QP 45	QP 44	QP 38
Hall2 V_{EWOC}	QP 46	QP 43	QP 39	QP 34
Hall2 D_{EWOC}	QP 48	QP 44	QP 40	QP 36
Street V_{full}	QP 48	QP 46	QP 42	QP 38
Street D_{full}	QP 44	QP 44	QP 38	QP 34
Street V_{EWOC}	QP 46	QP 44	QP 40	QP 37
Street D_{EWOC}	QP 44	QP 40	QP 36	QP 32

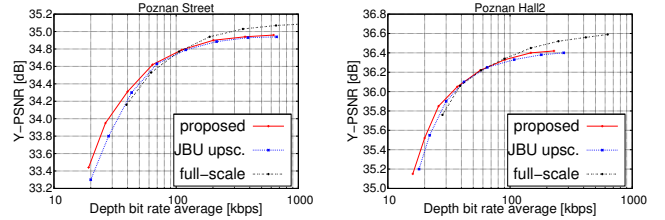


Fig. 3. Rate-distortion curves at different depth compressions.

and clean coding refresh (CDR) pictures to fulfill the MPEG CfP requirements for random access points every 0.5 seconds. All sequences were coded with the same settings, only varying spatial resolution and QP. No multiview video coding (MVC) was applied.

For both tests we synthesized from left and right view to the center view using the MPEG "View Synthesis Reference Software" (VSRS [15]). For objective evaluation we provided rate-distortion graphs for both tests. Distortion is represented as Y-PSNR, based on the error between the synthesized view and the true center view. For the first test we used the bit rate for one view (average between left and right), for the second test, the overall bit rate including texture and depth for both views.

IV. RESULTS

The rate-distortion graphs in Fig. 3 show the validity of our proposal as a depth map coding scheme. At bit rates below 105kps for 'Street' and 60kbps for 'Hall2', it outperforms full-scale depth map coding. For higher bit rates the distortion introduced by down-/upscaling is too large to allow competitive quality, as the similar trends for JBU upscaled depth and our proposal show. For lower bit rates, coding efficiency is increased by up to 16.5% for 'Street' and 18.5% for 'Hall2' compared to full-scale coded depth, including both the effects of depth down-/upscaling and adaptive pre-filtering. The comparison between JBU and our proposal shows an 10% increase in coding efficiency for 'Street' and 12% for 'Hall2' at the same points, stating the advantage of adaptive pre-filtering. These numbers show also how scene content influences coding results. 'Hall2' with its large uniform areas gets even better results than the complex depth structure of 'Street'.

In the second test procedure, the gained depth coding efficiency was used to increase the texture quality. The results are shown in Fig. 4. We replicate the results from the first

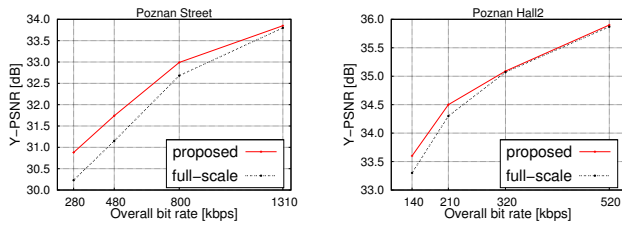


Fig. 4. Rate-distortion curves at MPEG CfP bit rate anchors [12].

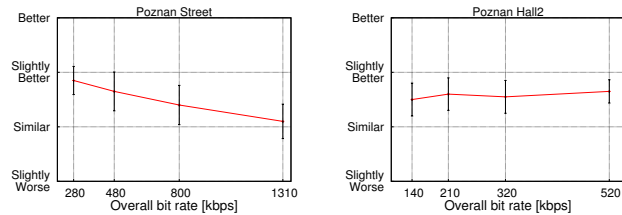


Fig. 5. Mean opinion score and 95% confidence intervals comparing sequences from Fig. 4.

test procedure, showing even higher efficiency at the three lower bit rate points with up to 0.6dB increase in Y-PSNR. At the higher bit rates we now obtain slightly better or similar objective quality, since the negative effects shown in Fig. 3 were compensated with higher texture quality. One danger using texture information for depth upscaling is the effect the compression has on the edge detection. Our approach handles this well, as the good results for low overall bit rates show.

These objective results were verified in our subjective test. For test sequence 'Hall2' all subjects picked similar or better quality for view synthesis with our depth map coding approach. And the same for the two lower bit rate points of 'Street'. As Fig. 5 shows, there is no significant preference for the highest bit rate point of 'Street'. Still the overall results show a statistical significant preference of the proposed method in 7 out of 8 test cases. We expect to achieve even larger gains in a stereo viewing scenario, where binocular fusion will improve the experienced quality.

V. CONCLUSIONS

In this paper we have introduced a novel depth map coding scheme for the MVD format, using depth up-/downscaling and adaptive filtering. Unlike similar approaches we extend the concept of bilateral information for upscaling to include a non-uniform lowpass filtering before downscaling to increase HEVC coding efficiency. Depth map sequences from an MVD signal are downscaled by a factor of 2, compressed with HEVC and upscaled with EWOC, using texture edge information. Objective results showed up to 18.5% gain in coding efficiency for lower bit rates. Compared to a similar depth down-/upscaling approach, our adaptive pre-filtering brought around 10% increase in coding efficiency. At higher bit rates the distortions introduced by down-/upscaling were too prominent

to achieve competitive results, when using a constant texture bit rate. If the reduction in depth bit rate was used to increase the texture bit rate budget, we achieved similar results at higher bit rates. Subjective evaluation showed a statistical significant preference for the proposed method in 87.5% of the presented test cases, compared to full-scale HEVC depth map coding.

Future research will include more detailed subjective testing, especially for stereo viewing. It can be assumed that this will lead to a larger gain in perceived quality, since binocular fusion with the less compressed texture will likely compensate for the synthesized view. Further points that need to be addressed are complexity and processing time, quality of edge detection and possible effects of inter-view coding.

ACKNOWLEDGMENT

This work has been supported by grant 2009/0264 of the KK Foundation, Sweden, by grant 00156702 of the EU European Regional Development Fund, Mellersta Norrland, Sweden, and by grant 00155148 of Länsstyrelsen Västernorrland, Sweden.

REFERENCES

- [1] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Image Commun.*, vol. 22, pp. 217–234, February 2007.
- [2] JCT-VC, "Test model under consideration," JCTVC-A205, April 2010, Dresden, Germany.
- [3] C. Fehn, P. Kauff, M. O. D. Beeck, F. Ernst, W. Ijsselstein, M. Pollefeys, L. V. Gool, E. Ofek, and I. Sexton, "An evolutionary and optimised approach on 3D-TV," in *Proceedings of International Broadcast Conference (IBC)*, 2002, pp. 357–365.
- [4] A. Smolic, K. Müller, N. Stefanoski, J. Ostermann, A. Gotchev, G. Akar, G. Triantafyllidis, and A. Koz, "Coding algorithms for 3DTV - a survey," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1606–1621, 2007.
- [5] K. Klimaszewski, K. Wegner, and M. Domański, "Influence of views and depth compression onto quality of synthesized views," ISO/IEC JTC1/SC29/WG11 MPEG2009/M16758, July 2009, London, UK.
- [6] E. Ekmekcioglu, M. Mrak, S. Worrall, and A. Kondoz, "Utilisation of edge adaptive upscaling in compression of depth map videos for enhanced free-viewpoint rendering," in *Image Processing (ICIP)*, 2009 16th IEEE International Conference on, November 2009, pp. 733–736.
- [7] M. Wildeboer, T. Yendo, M. Tehrani, T. Fujii, and M. Tanimoto, "Color based depth up-sampling for depth compression," in *Picture Coding Symposium (PCS)*, 2010.
- [8] Y. Li and L. Sun, "A novel upscaling scheme for depth map compression in 3DTV system," in *Picture Coding Symposium (PCS)*, 2010, December 2010, pp. 186–189.
- [9] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upscaling," *ACM Transactions on Graphics*, vol. 26, no. 3, 2007.
- [10] S. Schwarz, M. Sjöström, and R. Olsson, "Depth map upscaling through edge weighted optimization," in *Proceedings of the SPIE, vol 8290: Conference on 3D Image Processing (3DIP) and Applications*, 2012.
- [11] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K. Wegner, "Poznań multiview video test sequences and camera parameters," ISO/IEC JTC1/SC29/WG11 MPEG2009/M17050, October 2009, Xian, China.
- [12] "Call for proposals on 3D video coding technology," ISO/IEC JTC1/SC29/WG11 MPEG2011/N12036, March 2011, Geneva, Switzerland.
- [13] ITU, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep. ITU-R BT.500-12, 2009.
- [14] "HM4: High efficiency video coding HEVC test model 4." [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-4.0/
- [15] "Report on experimental framework for 3D video coding," ISO/IEC JTC1/SC29/WG11 MPEG2010/N11631, October 2010, Guangzhou, China.