Li, Y., Sjöström, M., Jennehag, U., & Olsson, R., "A Scalable Coding Approach for High Quality Depth Image Compression", in *Proc. 3DTV conference, Oct. 2012*.

http://dx.doi.org/

A SCALABLE CODING APPROACH FOR HIGH QUALITY DEPTH IMAGE COMPRESSION

Yun Li, Mårten Sjöström, Ulf Jennehag, Roger Olsson

Dept. of Information Technology and Media, Mid Sweden University
SE-851 70 Sundsvall Sweden

ABSTRACT

The distortion by using traditional video encoders (e.g. H.264) on the depth discontinuity can introduce disturbing effects on the synthesized view. The proposed scheme aims at preserving the most significant depth transition for a better view synthesis. Furthermore, it has a scalable structure. The scheme extracts edge contours from a depth image and represents them by chain code. The chain code and the sampled depth values on each side of the edge contour are encoded by differential and arithmetic coding. The depth image is reconstructed by diffusion of edge samples and uniform sub-samples from the low quality depth image. At low bit rates, the proposed scheme outperforms HEVC intra at the edges in the synthesized views, which correspond to the significant discontinuities in the depth image. The overall quality is also better with the proposed scheme at low bit rates for contents with distinct depth transition.

Index Terms — Depth image coding, 3DTV, View synthesis

1. INTRODUCTION

Intermediate views can be rendered from multiple textures plus depth by using Multi-view Video plus Depth (MVD) format, saving bandwidth used for transmission compared to simulcast. The depth image can be compressed by state of the art standardized compression techniques, such as JPEG 2000, H.264/AVC [1] and the latest standard to be HEVC [2]. However, traditional encoders like H.264 blur depth discontinuities, which cause disturbing effects around edges in the synthesized views [3]. This brings the question if an alternative encoding scheme may achieve better quality at edges in the synthesized virtual views, and competitive quality in overall.

Compressing a depth image with HEVC intra encoder introduces more errors at larger discontinuities than in smoother areas. To achieve acceptable quality of the synthesized view from such a depth image, it needs to be compressed with low quantization parameter (QP) values, which implies a large bit rate. The problem of edge blurring has been addressed by applying edge-based methods. Paper [4] describes algorithm that losslessly codes the contours and the coefficients of polynomials that approximate the smooth areas bounded by edges. Edge-based image compression was proposed for cartoon images for which the contours and the pixels on both sides are coded by JBIG and PAQ, respectively; the images are reconstructed by homogeneous diffusion [5]. Edge-based compression was applied to depth image in [6], which utilizes JBIG to encode the contours, and DPCM to compress both the pixels around the contours as well as the uniform sparse sampling points of the depth image.

In this paper, we propose an intra frame compression scheme for high quality depth images. The scheme is based on lossless coding of edge-contours, uniform sparse sampling and smooth inpainting. The goals are to retain the inherent distribution of depth images with good quality, and to investigate the quality of rendered views in comparison to state of the art compression methods. We define the significance of depth discontinuities (edges) as the magnitude of the first order derivative of a depth image.

The proposed scheme is an independent work, which, however, exhibits certain similarities with the work in [6]. The works differ in several important aspects: The novelties of the paper are the following: (1) we introduce a pre-processing of edges to reduce the depth incoherence on both sides of an edge contour, which leads to an accurate depth image reconstruction. (2) The scheme is scalable with respect to quality at edges with decreasing significance as a consequence to the proposed structure. (3) We consider an evaluation method that reflects the quality at significant changes in depth in a rendered virtual view, as a complement to more traditional evaluation methods. The sequel of the paper is organized as follows. We outline our method and test set-up in Section 2 and 3, respectively. The results are presented in Section 4. Section 5 concludes our work.

2. PROPOSED METHOD

The proposed coding scheme is shown in Figure 1. It extracts edge contours from a depth image, after which an edge pre-processing is applied to render more coherent depth values on each side of the contour. Chain codes representing the edge contours and uniform sub-sampling points on both sides of the edge contours are then compressed by differential and arithmetic coding. In parallel, a very low quality depth image is encoded by HEVC intra in order to obtain uniform sparse sampling points on the receiver side. During decoding and reconstruction, the edge contours and the sub-sampled pixels of both sides are recovered, from which full edge information on either side of the contours are interpolated. The depth image is finally reconstructed by diffusion using a second order partial differential equation (PDE). Each block in Figure 1 is described below.

2.1. Encoding

Edge detection extracts the edges from the depth image. The scheme uses the Canny edge detector [7], which applies a high and a low threshold to identify edges: above the high threshold, all points are considered an edge, above the low threshold all points entirely connected are parts of an edge. Hence, insignificant edges can be removed by decreasing the detection sensitivity, i.e. by increasing the Canny edge thresholds by a factor T_e. The proposed

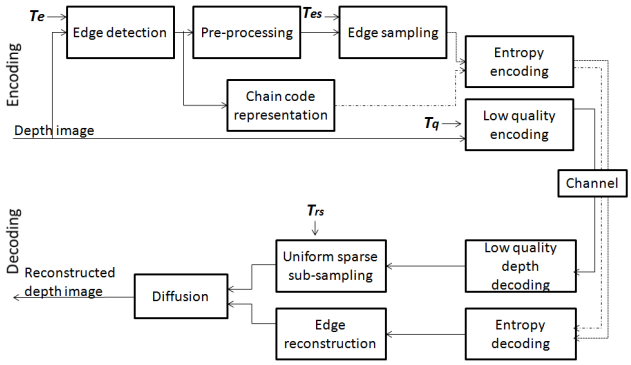


Figure 1. Coding system.

algorithm further removes edge contours that are less than 20 pixels long to ensure a more efficient compression. The total set of detected contours are denoted E_c.

Chain code representation: The edge contours are represented by Freeman chain code with 8 directions: from "right", "down right", "down" to "up right" in a clockwise manner. Each contour consists of the coordinates of the starting location (x,y), length (L) of the contour (i.e. the number of elements) and the direction of each element.

Pre-processing is introduced to make pixels along each side of the edge contour more homogeneous and reduce the sampling error in the next step. For this purpose, an edge mask E_m is generated by morphological dilation from

$$E_m = E_c \oplus S3 \tag{1}$$

with a square structure element $S3$ of size 3, where $\oplus$ is morphological dilation. A larger size than 3 is not chosen to avoid very thin parts of an object being removed. Figure 2 illustrates the pre-processing step. The pixels on each side of the edge contour can

(a) (b) (c)

Figure 2. Edge pre-processing: (a) a portion of the depth image before processing, the white line is the edge detected, (b) pixels removed by the mask in the black part, and (c) after processing.

contain heterogeneous pixels due to imperfections in the depth image, illustrated in (a). The pre-processing then removes the pixels within the edge mask as depicted in (b). The pixels within mask E_m near the contour E_c is assumed to be similar to its adjacent horizontal and vertical pixels:

$$\begin{aligned} &f(x,y) = f(x+1,y) = f(x,y+1), \\ &\text{such that } (x,y),(x+1,y),(x,y+1) \notin E_c. \end{aligned} \tag{2}$$

The image in (c) is obtained by solving the equations (2) in a least square error sense.

Edge sampling sub-samples pixels uniformly along both sides of the edge contours by a factor of T_{es}. The scheme uses sampled pixels on both sides of the edge contours; pixels on the contour are not encoded.

Entropy encoding: Contour directions as well as sub-sampled pixels on both sides of the edge contours are firstly differentiated. Then arithmetic coding is applied to this differentiated information and to the edge contour location and length (x, y, l).

Low quality encoding: The reason for transferring a low quality depth image encoding is to extract uniform sub-sampled depth image information for the reconstruction on the receiver side. The scheme encodes the full depth image with a high QP value and then sub-samples the low quality depth image on the receiver side.

2.2. Decoding

Entropy decoding: During decompression, the edge information including the edge contours and the sparse sampling pixels on both sides of the edge contours are decoded.

*Edge reconstruction:*The edge contours are reconstructed from the chain code, edge sub-sampled pixels are interpolated and placed along both sides of the edge contour. The scheme takes the pixels from nearest location on the right side of the edge contour to fill the pixels on the edge contour itself.

Low quality depth image decoding: The low quality depth image is decoded.

Uniform sparse sub-sampling: Uniform sparse sampling points of a sub-sampling factor T_{rs} and the full image border are extracted from the low quality depth image. Sparse points are discarded within five pixels width of the edge contours, so that they don't affect the reconstructed image quality around edges, as they can be heavily distorted due to the low quality encoding. The rationales of using a full compressed low quality depth image are the following: (1) For scalability reasons, as explained in Section 2.3. If all edge information is lost, this low quality full depth image can directly be used for view synthesis. (2) For a adjustable reconstruction in the *Diffusion* step, the uniform sparse sampling factor T_{rs} can be adjusted. The factor governs the diffusion complexity (the number of equations to be solved) versus smoothness.

*Diffusion:*With the information of reconstructed edges, the uniform sparse points and the image border, the scheme uses the Laplace equation,

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0, \tag{3}$$

in our implementation approximated by

$$\begin{aligned} f(x,y) \quad = \quad & [f(x-1,y) + f(x+1,y) + \\ & f(x,y-1) + f(x,y+1)]/4 \end{aligned} \tag{4}$$

as an interpolation method to diffuse the entire depth image. The linear equations are solved by the least mean square error method.

2.3. Scalability

We demonstrate here that the proposed scheme is scalable. Scalability implies in this context that the more information the decoder receives, the better the image reconstruction quality is. In the proposed scheme, the priority of the data information is related to the significance of the edges. The edges are detected by Canny edge threshold C_n, it is changed by a factor $T_e = n$, where $n = 1, 2...N$, and N is the index for the highest threshold C_N. The edge detector will produce the most significant edges with the highest threshold. Assumed that the edges extracted by Canny edge detector is $E_c(n) = f_c(C_n)$, where f_c is the function that outputs edges by Canny edge detector according to the threshold C_n, and the edges are represented by $E_c(n)$; a set e of edges can be built with $e(n) = E_c(n) - E_c(n+1)$. The significance of the information in descending order lists as $E_c(N), e(N-1), e(N-$

$2), ..., e(1)$. Once the more significant set of edges and the associated sampling pixels on both sides of the edge contours are available, the better quality depth image can be reconstructed. If the complete edge set is lost, the decoded full resolution low quality depth image is still available for view synthesis.

3. TEST ARRANGEMENT AND EVALUATION CRITERIA

In this test, only the depth images were compressed. The decompressed depth image and the original texture were used to synthesize virtual views. VSRS [8] version 3.5 was utilized for the view synthesis. We evaluated the quality of the synthesized virtual view with respect to a reference view, which was synthesized from the original depth image and the original texture.

Two test images consisting of multi-view images plus depth information were selected for the test: the first frame of the sequence Poznan Street [9] and the 239th frame of Lovebird1 from Electronics and Telecommunications Research Institute (ETRI). The rationale for choosing these data sets is that the former is a complicated scene with gradual depth transition at background and relatively large depth discontinuities for the foreground objects; the latter contains a very large depth discontinuity for the foreground object while background depth is relatively smooth with fewer edges. The depth images of both data sets have relatively accurate edges in the sense that the edges coincide well between the depth image and the associated texture. The virtual views were synthesized at equal distance from two given views: for the first test image, the view was synthesized from camera 3 and 5 at camera position 4, and for the second test image, a view was synthesized at camera position 6 from camera 4 and 8.

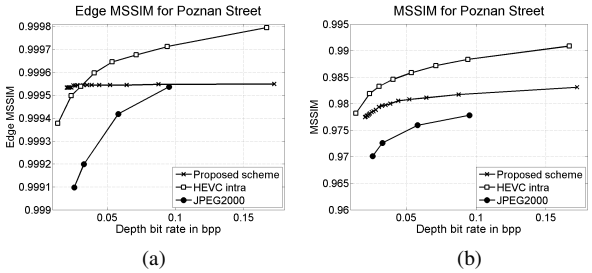
The proposed scheme was implemented using Matlab 2010b. HEVC intra with QP 41 was selected for acquiring the low quality depth image in our test. We compared the proposed scheme with JPEG2000 and with HEVC intra frame coding using HEVC version 6.5 with QP 21, 25, 27, 29, 31, 33, 35 and 39. Only the depth image was compressed. The decompressed depth image and the original texture were used to render virtual views in the same manner as for the proposed scheme.

The scheme contains a certain number of parameters, which must be adjusted to give an optimized quality with respect to bit rate. This optimization is out of the scope of this paper. In the test, we used values: $T_{es} = 30$ and $T_{rs} = 8$. We changed the edge threshold factor T_e from 1.5 to 10 with a step of 0.5 for both depth images from Poznan Street and Lovebird to produce bit rates approximately within the range of that produced by the reference compression method. The edge thresholds used in the tests were calculated by multiplying T_e by the default thresholds of the Canny edge detector in Maltab, which are obtained by running the Canny edge detection with default parameters. The default thresholds were [0.0063, 0.0156] for Poznan Street and [0.0125, 0.0313] for Lovebird, respectively.

The evaluation was partly performed by using structural similarity (SSIM) [10] index on the luminance components of the full synthesized image. The rate-distortion graphs for the test are based on MSSIM values versus bit rates of the compressed depth images from two views. In order to assess the results in the neighborhood of edges, we use an evaluation method we call Edge MSSIM. The reason for using this method is that edges at the synthesized view corresponding to the significant depth discontinuities are generally more distorted by the synthesis process and so they draw the attention of the observer; objects at significant depth changes are often what draws the attention of the viewer. These edges will therefore influence the perceived quality. The Edge MSSIM computes the MSSIM on the full processed synthesized view in which areas other than the edge mask are cleared to zero. The edge mask is generated by applying morphological dilation with a square structure element of size 8 to the significant edges. The edges are identified by applying the Canny edge detector on the given depth images at the position of the rendered virtual views, i.e. view 4 for Poznan Street and view 6 for Love bird. The thresholds are 18 times the default thresholds of the Canny edge detector. The edges resulted from these thresholds are the edges of the front most objects with the most significant depth discontinuities. Figure 3 illustrates the edge mask for Poznan Street.

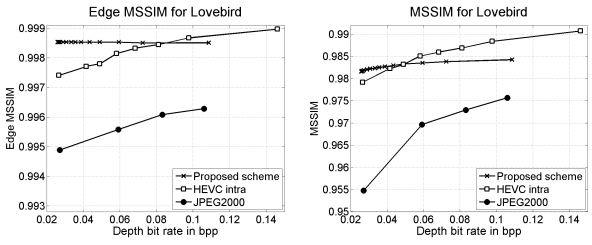
Figure 3. The edge mask used by Edge MSSIM for Poznan Street.

4. RESULTS AND ANALYSIS



(a) (b)

Figure 4. Evaluation results for the synthesized view (Poznan Street). (a) Edge MSSIM. (b) MSSIM.



(a) (b)

Figure 5. Evaluation results for the synthesized view (Lovebird1). (a) Edge MSSIM. (b) MSSIM.

The results for Poznan Street calculated by Edge MSSIM (Figure 4(a)) show that the proposed scheme performs better with respect to the given significant depth discontinuities in the synthesized view for bit rates below 0.0302 bits per pixel (bpp). These bit rates were produced by setting the edge threshold multiplication factor T_e above 5.5. The figure further shows that edge quality appears to be independent of bpp. This is partly because the edges assessed by Edge MSSIM were always included by the proposed scheme for decoding and reconstruction at all tested bit rates. This behavior is a consequence of how the Edge MSSIM was defined, and further investigations is required to say if this is also valid for a

perceptual evaluation. Also for the image from Lovebird, the proposed scheme outperforms HEVC for bit rates below 0.085 bpp when considering Edge MSSIM, see Figure 5(a).

The result for the full image MSSIM in Figure 4(b) illustrates that HEVC outperforms the proposed scheme. Nonetheless, the proposed scheme performs better than HEVC for the whole image from Lovebird at bit rates below 0.049 bpp, as illustrated in Figure 5(b). The image from Lovebird has much larger depth discontinuities, whereas The Poznan Street contains more edges and has gradual depth transitions. This would then imply that the proposed scheme is advantageous not only at the edges but also on the full image of the synthesized view, when the depth image contains fewer edges and has prominent depth discontinuities.

A visual inspection of the results reveals that the proposed scheme produces less distortion in the synthesized views at the significant discontinuities, not only for certain but for all the tested bit rates. This also implies a limitation of the objective metric for evaluating visual quality. Detail portions of the rendered images for Poznan Street are shown in Figure 6 corresponding to bit rate 0.0236 bpp using HEVC and the proposed scheme, respectively. In Figure 6(b), the blurring effect of the pole at the background is due to that the insignificant edges for the pole were not encoded at this bit rate.

(a) (b)

Figure 6. View synthesis results. (a) HEVC. (b) Proposed method.

An alternative way of more efficient coding to the low quality encoding for obtaining the uniform sub-sampling pixels of the full depth image can be taken by down-sampling and compressing at the encoder side. This alternative has the disadvantage that the full image border must be recreated by interpolation. We have tested such an implementation in the proposed coding system, it produced worse results for some parts of the image by our visual inspection. However, it is subjective to further research.

The proposed encoding scheme retains the inherent disposition of depth images, as it strives to maintain accurate discontinuities in the depth image in order to have truthful synthesis of virtual views. If such discontinuities are erroneous, e.g. not coinciding with texture, a smoother transition of the discontinuities may lead to less errors and a better quality of experience of the virtual view. In such a case, HEVC may be advantageous because it blurs the edges and strives for the best fidelity for the entire depth image.

5. CONCLUSIONS

We have proposed and implemented a scalable depth image coding scheme that retains the inherent distribution of depth images, as it accurately preserves discontinuities in the depth image. The scheme extracts the edge contours and represents them by chain code. The chain code and the uniform sub-sampling depth values along both sides of the edge contours are encoded by differential and arithmetic coding. The depth images are reconstructed by diffusing the edge information and the uniform sparse sampling points using the Laplace equation. The sparse sampling information is obtained from a low quality depth image compressed by HEVC intra with high QP. We theoretically demonstrated that the proposed scheme is scalable.

The test results showed that, at the lower bit rates, the proposed scheme can produce higher quality at edges in synthesized views than HEVC intra encoding, which is a consequence of the edges in the synthesized view corresponding to significant depth discontinuities in the depth image. In the case when the depth image contains large depth discontinuities, the proposed scheme results in a higher overall quality for the full synthesized view than HEVC at the lower bit rates.

The investigation showed that the proposed scheme may very well be an alternative to encode high quality depth images with accurate depth information.

Future work comprises to verify the obtained results in subjective test, and to compress contours and sub-sampling pixels more efficiently, as well as to include the texture as an aid into the depth map compression.

6. ACKNOWLEDGMENT

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

7. REFERENCES

[1] ITU-T, "Advanced video coding for generic audiovisual services," *Recommendation ITU-T H.264*, Jan. 2012.

[2] JCT-VC, "WD1: Working Draft 1 of High-Efficiency Video Coding," *JCTVC-C403*, Oct. 2010, Guangzhou, China.

[3] S. Liu, P. Lai, D. Tian, C. Gomila, and C.W. Chen, "Joint trilateral filtering for depth map compression," *Image Processing*, vol. 7744, pp. 77440F1–10, 2010.

[4] F. Jager, "Contour-based segmentation and coding for depth map compression," *2011 Visual Communications and Image Processing (VCIP)*, pp. 1–4, Nov. 2011.

[5] M. Mainberger and J. Weickert, "Edge-Based Image Compression with Homogeneous Diffusion," *American Journal of Psychology*, pp. 1–8.

[6] J. Gautier, O. Meur, and C. Guillemot, "Efficient Depth Map Compression based on Lossless Edge Coding and Diffusion," pp. 81–84, 2012.

[7] J. Canny, "A computational approach to edge detection.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 8, no. 6, pp. 679–98, June 1986.

[8] "Report on experimental framework for 3D video coding," *ISO/IEC JTC1/SC29/WG11 MPEG2010/N11631*, Oct. 2010, Guangzhou, China.

[9] M. Domaski, T. Grajek, K. Klimaszewski, and M. Kurc, "Pozna Multiview Video Test Sequences and Camera Parameters," *ISO/IEC JTC1/SC29/WG11*, 2009.

[10] W. Zhou, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600 –612, April 2004.