INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2014/M34298
July 2014, Sapporo, Japan

Source **Poznan University of Technology,**
Chair of Multimedia Telecommunications and Microelectronics
Status **Contribution**
Title **Fast View Synthesis through Depth Simplification**
Author Krzysztof Wegner (kwegner@multimedia.edu.pl)
Olgierd Stankiewicz
Marek Domański

1 Introduction

This documents presents an approach to accelerate view synthesis algorithm by use of simplified model of the depth data. The depth map is approximated in square blocks with flat plane models. Thanks to that, the model based view synthesis can be performed in blocks which lowers required number of point transformations during view synthesis.

2 Fast View Synthesis

The main idea of the proposed fast view synthesis lies in point transformation simplification.

Depth maps are the images where every pixel represents distance to the object in the scene. At the same time, depth maps for natural sequences are very smooth, rarely divided, so dense distance representation is redundant. Moreover, during view synthesis process, such depth map representation causes unnecessary operation while projection.

In our approach, smooth areas of depth are be modeled by flat planes. It is done in blocks or variant size. Each block is described by its 4 corners. Such a representation significantly reduce complexity of view synthesis process. Instead of pixel by pixel view transformation, model based transformation is performed. For a given block of size $N \times N$ pixels, only 4 corners of the plane model have to be transformed instead N^2 pixels (fig. 1). Such an approach leads to speed-up of the block transformation process, which can be described by acceleration factor:

$$acceleration = \frac{N^2}{4} \quad . \tag{e1}$$

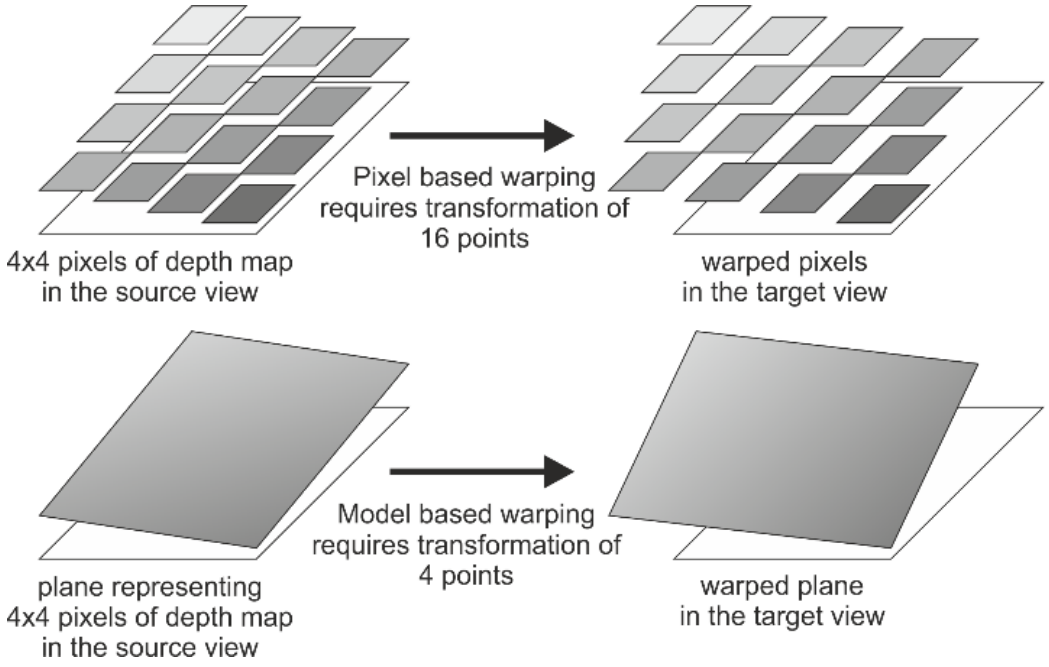

Figure 1. The idea of the model based view synthesis acceleration.

3 Adaptive depth map simplification

3.1 Depth map partitioning

In proposed method each depth map is divided into non-overlapping hierarchy of blocks. At top level, the block size is 64x64. Each such block B can be adaptively partitioned into smaller sub-blocks B_i of various sizes N_i in a quad-tree manner. Therefore, each block B can be composed of a set of its sub-blocks B_i which cover surface of entire block (fig 2).

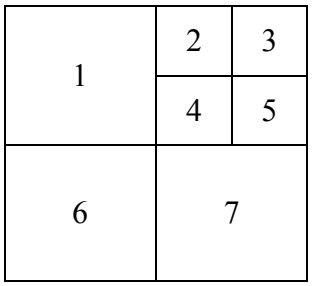

Figure 2. Exampled block partitioning onto sub-blocks 1 to 7.

Partitioning is performed in a such way, to provide maximum transformation/view synthesis acceleration while maintained high quality view synthesis. To do so partitioning process is controlled by Lagrangian optimization:

$$fitness = \lambda \cdot error + acceleration \quad , \tag{e2}$$

where $error$ is a depth maps block modeling error, and $acceleration$ is amount of transformation speed-up provided by the block model. λ is a Lagrangian coefficient which allows control preference between view synthesis acceleration and quality of the view synthesis.

3.2 Subblock model

Each subblock of the block in quad-tree is approximated independently by a simple plane model defined by the following equation:

$$z'(x,y) = A \cdot x + B \cdot y + C \quad . \tag{e3}$$

where $z'(x,y)$ is a depth value for pixel within the approximated sub-block in coordinates (x,y) calculated based on the applied model. A, B, C defines plane model itself, and are constant for given sub-block in a quad-tree.
Optimal plan model in a sense for least-square error is defined by following parameters:

$$A = \frac{12S_{xz} - 6S_z(N-1)}{N^2(N+1)(N-1)} \quad , \quad S_{xz} = \sum_{x=0}^{N-1} x \sum_{y=0}^{N-1} z(x,y) \quad ,$$

$$B = \frac{12S_{yz} - 6S_z(N-1)}{N^2(N+1)(N-1)} \quad , \quad S_{yz} = \sum_{y=0}^{N-1} y \sum_{x=0}^{N-1} z(x,y) \quad , \tag{e4}$$

$$C = \frac{-6S_{xz} - 6S_{yz} + S_z(7N-5)}{N^2(N+1)} \quad , \quad S_z = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} z(x,y) \quad .$$

Where N is subblock size in pixels, and $z(x,y)$ is a depth value at coordinates (x,y) in a sub-block being currently modeled.

Simplification of depth in sub-blocks by plane approximation leads to some amount of depth quality degradation. Level of introduced error can be expressed as:

$$error = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left(z'(x,y) - z(x,y)\right)^2 \quad . \tag{e5}$$

3.3 Block acceleration and approximation error

Each sub-block B_i is modeled by a plane model defined by parameters A_i, B_i, C_i. Due to plane modeling each sub-block is approximated with some $error_i$.

Thus entire block B is approximated with an error which is a sum for the sub-block errors:

$$error = \sum_{i \in B} error_i \quad . \tag{e5}$$

Due to sub-block modelling, instead transforming each $N_i \times N_i$ sub-block pixels, pixel by pixel, which requires $\sum_{i \in B} N_i^2 = N^2$ point transformations in total for a whole block, for each sub-block, only 4 plane corners need to be transformed, which requires only $\sum_{i \in B} 4$ point transformations in total for a whole block. Thus block transformation acceleration provided by the proposal can be defined as:

$$acceleration = \frac{N^2}{\sum_{i \in B} 4} \quad . \tag{e5}$$

Figure 3 present exemplary simplified depth maps with applied partitioning scheme.

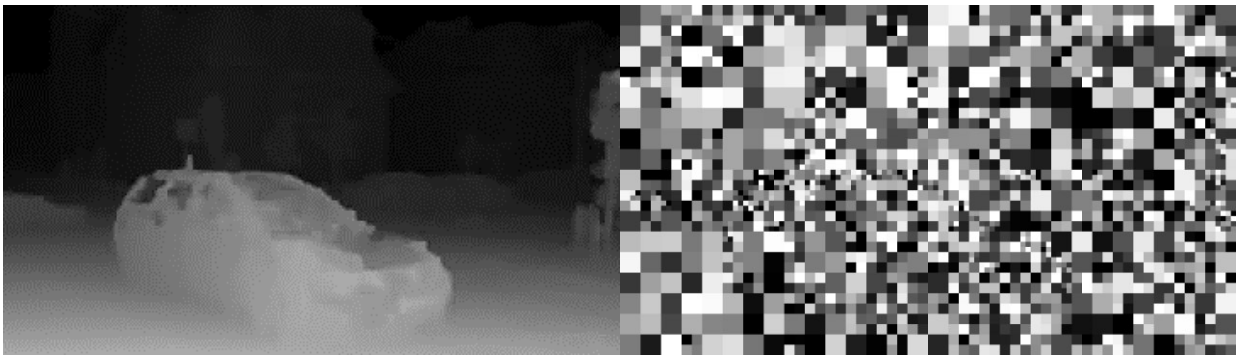

Figure 3. An exemplary depth frame of Poznan Street sequence [1], partitioned into a blocks of various size (right) and modeled (left). Partitioning has been illustrated as blocks uniformly shaded with random gray level.

4 Evaluation

In order to evaluate the proposed algorithm and asses real-world transformation acceleration , several experiments have been conducted.

Set of commonly used 3D video sequences with high quality depth maps recommended by ISO/IEC MPEG have been used (fig 4). Each sequence consists of 3 videos and 3 depth maps, along with camera parameters.

Figure 4. 3D test sequences with depth maps used in experiments, from top-left: Poznan Street, Poznan Hall, [1, 2], Undo Dancer [3], GT Fly [4], Kendo, Balloons [5] and Newspaper [6].

Each depth map of each sequence have been simplified and approximated by the proposed algorithm. Simplified depth maps and unprocessed original video have been used to synthesize 6 intermediate virtual views equally distributed in positions between position of the input views.

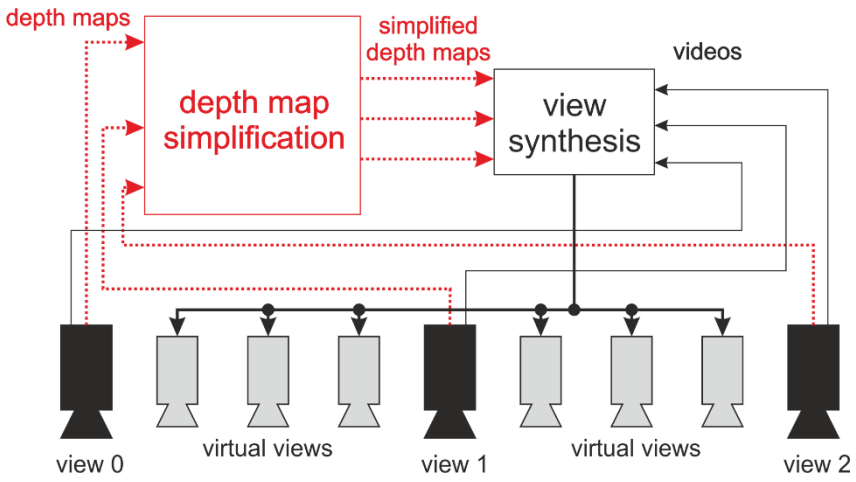

Figure f5. Scheme of the evaluation done.

Average quality measures have been gathered, in terms of luminance PSNR of those intermediate virtual view with respect to virtual view at the same spatial position synthesized based on unprocessed data. Experiments have been conducted for wide range of lambda coefficient with controls amount of acceleration versus quality of the modelled depth.

Another series of experiments, which have been conducted, relate to depth map coding. Depth maps compression influence view synthesis performance because some information are lost due to compression. Such a decompressed depth maps should be easier to approximate.
Those experiments have been performed with use of state of the are 3D video compression technology namely 3D-HEVC [7] developed by Join 3D Video Coding Team (JCT-3V). Each test sequences was encoded with 4 different rate points, as recommended by JCT-3V for compression performance evaluation [8]. Following QP/QD pairs have been used for video and depth maps: (25,34), (30,39), (35,42), (40,45). Decompressed depth maps have been simplified with used of proposed algorithm and virtual view synthesis has been performed.

5 Results

Figure 5 shows quality degradation of the virtual views synthesized based on simplified depth map. Even ×16 times acceleration provides high virtual view quality (over 40dB). Further accelerating view synthesis comes at a price of slight quality degradation but still provide good virtual view quality (near 35-40dB).

When decompressed depth map are being simplified wide range of acceleration up to ×16, ×32 provide negligible quality loses (fig 6). It means that at the same time it is possible to synthesize not one or two virtual views but up to 32 which make proposed approach very interesting for autostreoscopic displays where many view needs to be rendered at the same time.

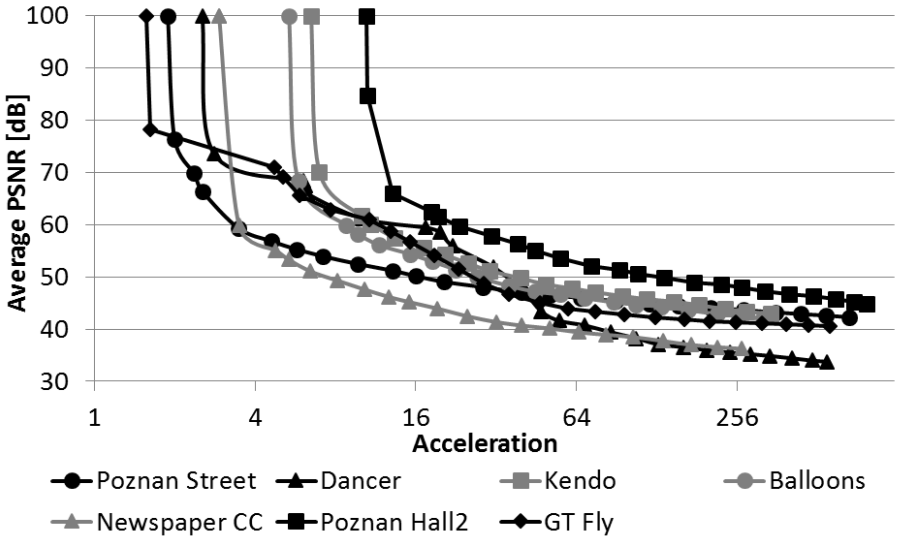

Figure f5. Average quality of virtual views versus transformation acceleration.

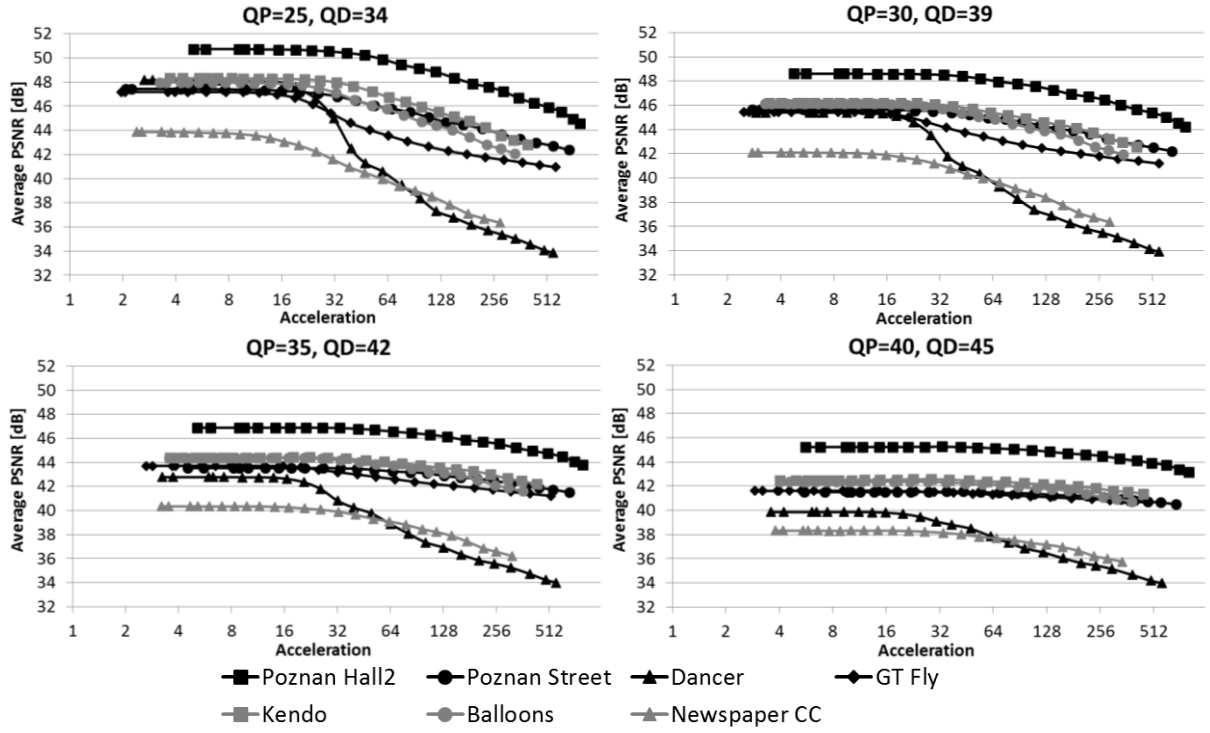

Figure f6. Average quality of virtual views synthesized base on videos and depths compresed by 3D-HEVC with various QP/QD pairs versus acceleration.

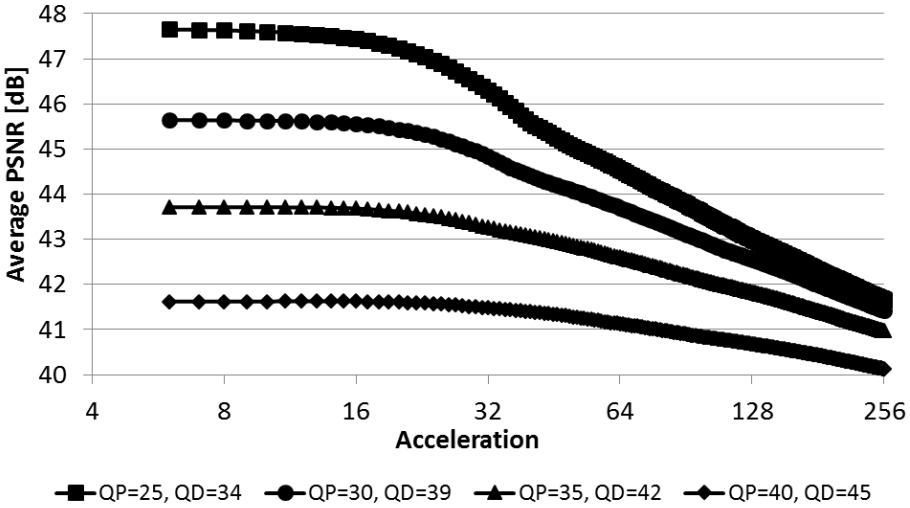

Figure f7. Comparison of view synthesis quality in terms of PSNR averaged over all of the sequences versus acceleration.

6 Conclusion

A novel approach to depth map simplification and view synthesis acceleration has been presented. As it has been shown, both theoretically and empirically, the proposed algorithm leads to ×16 to ×256 acceleration (speed-up) with negligible quality loss. Level of acceleration and quality degradation can be smoothly controlled by Lagrangian coefficient.

7 Acknowledgments

Research project was supported by National Science Centre, Poland according to the decision DEC-2012/05/N/ST6/03378.

8 References

[1] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, „Poznań Multiview Video Test Sequences and Camera Parameters ”ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, Xian, Chiny, Październik 2009.

[2] O. Stankiewicz, K. Wegner, M. Domanski, “First version of depth maps for Poznan 3D/FTV test sequences”, ISO/IEC JTC1/SC29/WG11, MPEG2010/M17176, Kyoto, Japan, January 2010.

[3] D. Rusanovskyy, P. Aflaki, and M. M. Hannuksela, “Undo dancer 3DV sequence for purposes of 3DV standardization,” ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Tech. Rec. M20028, Mar. 2011.

[4] J. Zhang, R. Li, H. Li, D. Rusanovskyy, and M. M. Hannuksela, “Ghost town fly 3DV sequence for purposes of 3DV standardization,” ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Tech. Rec. M20027, Mar. 2011.

[5] M. Tanimoto, T. Fujii, M. P. Tehrani, M. Wildeboer, N. Fukushima, H. Furihata, “Moving Multiview Camera Test Sequences for MPEG-FTV“,ISO/IEC JTC1/SC29/WG11 M16922, Xian, China, October 2009.

[6] Y.-S. Ho, E.-K. Lee, C. Lee “Video Test Sequence and Camera Parameters” ISO/IEC MPEG M15419, Archamps, France, April 2008.

[7] G. Tech, K. Wegner, Y. Chen, S.Yea, “Test Model 8 of 3D-HEVC and MV-HEVC” JCT-3V of ITU-T SG 16 WP 3 and ISO/IEC JTC1/SC 29/WG 11, Doc. JCT3V-H1003, Valencia, ES, 2014.

[8] "Common Test Conditions of 3DV Core Experiments" Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Document: JCT3V-G1100, 7th Meeting: San José, USA, 11–17 Jan. 2014.