Pacific Graphics (2014)
J. Keyser, Y. J. Kim, and P. Wonka (Editors)

Saliency-driven Depth Compression for 3D Image Warping

Minjie Gu[1], Shanfeng Hu[1], Xiaochuan Wang[1], Xiaohui Liang[1], Xukun Shen[1] and Aihong Qin[2]

[1] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China
[2] Zhejiang Institute of Media and Communications, Hangzhou, China

Abstract

Current compression methods compress depth images by incorporating 2D features, which leads to a loss of the detail of the original 3D object in the recovered depth image. The main idea of this paper is to augment 2D features with 3D geometric information to preserve important regions of the depth image. Mesh saliency is used to represent the important regions of the 3D objects, and discontinuity edges are extracted to indicate the important regions of the depth image. We use mesh saliency to guide the adaptive random sampling to generate a random pixel sample of the depth image and then, combine this sample with the depth discontinuity edge to build the sparse depth representation. During the depth reconstruction, the depth image is recovered by using an up- and down-sampling schema with Gaussian bilateral filtering. The effectiveness of the proposed method is validated through 3D image warping applications. The visual and quantitative results show a significant improvement of the synthetic image quality compared with state-of-the-art depth compression methods.

Categories and Subject Descriptors (according to ACM CCS): I.3.m [Computer Graphics]: Miscellaneous—performance

1. Introduction

Depth-Image-Based Rendering (DIBR) techniques have been applied to many applications, including image-based remote rendering systems [KCTS01] [KTL*04] [PHE*11], image-based interactive 3D navigation [KKS*05], 3D video games [Smo11], 3DTV [PJO*09] and FTV [MFY*09], which use a depth image to warp the original camera position to nearby virtual camera positions to provide immersive 3D visual enjoyment. For all of these applications, the depth image and color image must be compressed, transmitted and stored. The compression of the color image follows several standards (e.g., JPEG-2000 and H.264), but there is no standard for the compression of the depth image. The depth image from the z-buffer is intrinsically different from the color image; it contains a depth value for each sample, which represents the current closest depth of all previously rendered triangles that overlap the sample [AMAM13].

Recently proposed depth compression methods can be classified into five types: transform-based encoding (TE) [Mar02], intra/inter-frame predictive encoding (IE) [SMW07] [FMY*13] [NMD13] [PHM*14], piecewise-linear function fitting (PT) [MF*06], mesh-based depth representation (MR) [CSS02] and sampling-based depth reconstruction (SR) [WYT*10] [PHE*11] [PHM*14] [NMD13]. These methods generally incorporate the depth image's 2D features to perform the compression. TE-based methods consider the depth image as a color image and use DCT transforms to perform the compression but do not use the depth image's inherent characteristics. IE-based methods consider the depth image sequence as a stream and exploit the intra-frame's macro block matching and the inter-frame's residual to reduce the data prior to transmission. In contrast, PT-based methods exploit the smooth characteristic of depth signals and use piecewise-constant and piecewise-linear functions to fit the depth signals using complex algorithms. MR-based methods regard the depth field as the height field and use the method of Lindstrom et al. [LKR*96] to simplify the triangular meshes in real time. SR-based methods sample the original depth image to perform the compression and use interpolation to fill holes to complete the depth reconstruction.

Although these methods exploit the depth image's 2D features to perform the compression, none incorporate the 3D geometric information to compress the depth image. To utilize the 3D information, this paper proposes a new method of geometry-dependent depth image compression using hybrid saliency sampling. We consider information from both the 3D space and the 2D depth image to acquire the importan-

DOI: 10.2312/pgs.20141258

t regions of the depth image. The main idea of our method is to combine the 3D object's mesh saliency and the depth discontinuity edge to build the sparse depth representation.

The contributions of this paper are as follows:

- We present a saliency-driven depth image compression method that incorporates the mesh saliency and the depth discontinuity edge to build the sparse depth representation.
- The proposed method is applied to 3D image warping. The visual and quantitative results show a significant improvement of the synthetic image quality compared with state-of-the-art depth compression methods.

After reviewing related work in Section 2, we describe our method in detail in Section 3. The visual evaluation, quantitative evaluation and computational complexity are then discussed in Section 4. Finally, we draw conclusions and discuss limitations of the method and future work in Section 5.

2. Related Work

Depth compression. Previous work on depth image coding has used a transform-based algorithm that was derived from JPEG-2000 [Mar02]. Perceptually, such a code generates ringing artifacts along edges that lead to fuzzy object borders in the synthesized image. To overcome this problem, the method presented by Morvan et al. [MF*06] exploits straight lines to separate boundaries to model depth signals using piecewise-constant and piecewise-linear functions.

Chai et al. [CSS02] propose a regular triangular mesh-based depth representation that is inspired by Lindstrom et al. [LKR*96], who proposed an algorithm for real-time, continuous level of detail rendering of digital terrain and other height fields. The method achieves moderate compression by encoding vertices in the mesh rendering order, but it has a low depth image DPCM coding efficiency.

Wildeboer et al. [WYT*10] present a depth compression method that contains a depth upsampling filter that uses the color image as prior information. The method is able to maintain clear object boundaries in the reconstructed depth images and uses a regular grid with a constant step size to sample the depth image.

Inspired by the push-pull mechanism described by Gortler et al. [GGSC96], Pajak et al. [PHE*11] use the depth edge to sample the original depth image. To avoid missing low-frequency changes, their technique also uniformly adds one sample every 32*32 pixels. This method results in rapid depth compression, but the precision of the recovered depth is insufficient for 3D image warping.

Saliency detection for meshes. Mesh saliency represents a measure of the regional importance for graphics meshes. Lee et al. [LVJ05] introduce the concept of mesh saliency in a scale-dependent manner using a center-surround operator on Gaussian-weighted mean curvatures. Inspired by low-level human visual system cues, Lee et al. [LVJ05] indicate that this definition of mesh saliency is able to capture what most would classify as visually interesting regions on a mesh. This definition is also applied to salient-guided mesh simplification and salient viewpoint selection. To utilize perception-based metrics to process 3D meshes, we combine the mesh saliency and the depth image's discontinuity edge to compress the depth. The method proposed in this paper uses the strategy of Lee et al. [LVJ05] to efficiently compute the model's mesh saliency in the preprocessing stage.

Random sampling. Quinn et al. [QLLM13] note that patterns and regularities are quickly detected by humans and should be avoided. Poisson disk sampling [Wei08] can avoid a regular grid structure [LD08] and can generate a blue noise property. Based on a spectrum analysis, the power is concentrated at high frequencies, which is similar to the ground truth produced by dart throwing [Coo86].

Cook [Coo86] uses dart throwing to produce a Poisson disk sampling. However, this process is too slow for applications that require a large number of samples. Wei [Wei08] demonstrates a parallel Poisson disk sampling algorithm that runs on a GPU to generate samples with a blue noise spectrum. The basic idea of this method is to uniformly draw samples from square grid cells and to perform this operation concurrently and independently for all grid cells that are sufficiently far apart.

3D image warping. McMillan [MJ97] proposes 3D image warping as a classical DIBR technology. This method can be applied to image-based remote rendering systems, image-based interactive 3D navigation, 3DTV, FTV and virtual reality. It uses a referenced viewpoint's depth to generate the desired viewpoint's content to provide an immersive 3D visual impact. Koller et al. [KTL*04] note that DIBR is suitable for sharing archives of 3D models while protecting the 3D geometry from unauthorized extraction. Mark et al. [MMB97] demonstrate a method of post-rendering 3D warping and indicate that the method can be applied to real-time remote display systems. Shade et al. [SGHS98] demonstrate layered depth images, which are views of a scene from a single input camera with multiple pixels along each line of sight. The method uses the McMillan warping algorithm to generate the desired viewpoint's content. In this paper, we use 3D image warping to evaluate our method, including the precision of the recovered depth and the visual impact.

3. Approach

Our approach includes three stages: mesh saliency precomputation (described in [LVJ05]), sparse depth representation construction (described in subsection 3.1) and depth reconstruction (described in subsection 3.2). In the preprocessing stage, the model's mesh saliency is precomputed

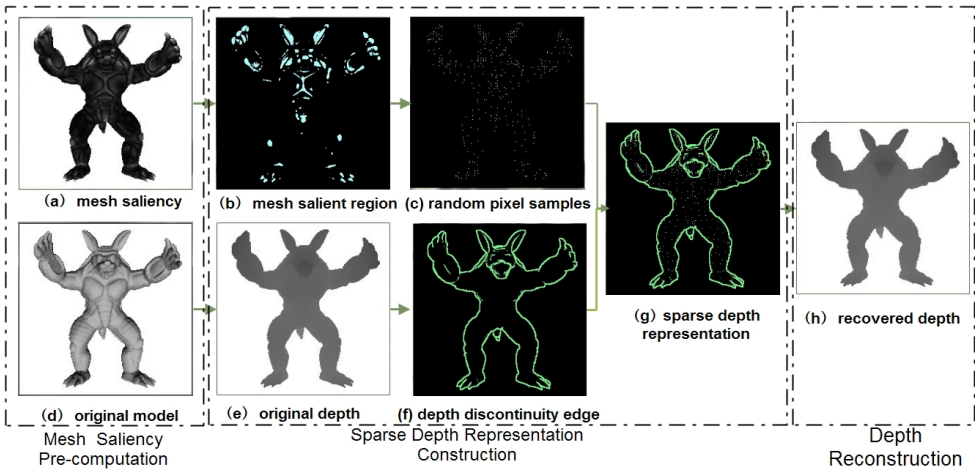

Figure 1: ***Overview of the approach****. The steps are shown from left to right and follow the arrows.*

using a center-surround operator on the Gaussian-weighted mean curvatures. Figure 1 (a) shows the output of the mesh saliency precomputation. The intensity of the model's surface represents the value of the mesh saliency. Figure 1 (d) shows the input of the mesh saliency precomputation.

Subsequently, our method uses the mesh's salient regions to guide the adaptive random sampling to generate the random pixel sample and subsequently combines the sample with the depth discontinuity edge pixel to build the sparse depth representation. Figure 1 (b) shows the 2D map of the mesh's salient region, which is generated by rendering the model's mesh saliency. Figure 1 (c) shows the random pixel sample, which is generated from the adaptive random sampling that is guided by the mesh's salient region. Figure 1 (e) shows the original depth image obtained from the depth buffer by rendering the original model, and Figure 1 (f) shows the depth discontinuity edge, which is detected by the Laplace operator on the original depth image. Figure 1 (g) shows the sparse depth representation, which is the output of the second stage.

Finally, the depth is reconstructed using an up- and down-sampling schema with Gaussian bilateral filtering. Figure 1 (h) shows the recovered depth, which is the result of the depth reconstruction.

3.1. Sparse Depth Representation Construction

Poisson disk sampling can generate an isotropic random sample with a blue noise property. We extends the procedure of Wei [Wei08] to generate the adaptive sampling. The algorithm runs on a GPU with the GLSL fragment shader, and we use the hash-based method to generate the random number as detailed in [TW08]. Alternatively, we can exploit the grid cell's coordinates and the time of each render pass (Render to Texture, RTT) as the random seed.

By controlling the sampling density, the proposed method exploits the mesh saliency to adjust the sparseness of the adaptive random sampling. Unlike the uniform sampling case, in which each sample has to maintain the same minimum distance r from every other sample, in adaptive random sampling, we use the function $r(.)$ over the sampling domain Ω and specify the minimum distance $r(s)$ for which the sample $s \in \Omega$ has to be maintained from other samples.

In this case, the domain Ω is the original depth image, and D is the mesh's salient region. We use the formulation presented below to describe the threshold function $r(s)$:

$$r(s) = \begin{cases} r_1 & \text{if } s \in \mathrm{D} \\ r_2 & \text{if } s \notin \mathrm{D} \end{cases} \quad (1)$$

The proposed method uses r_1 and r_2 to control the sampling density. We set $r_1 = 1$ and $r_2 = 6$; the user can set the other parameters as required, but r must be less than $2^{(scale-1)}$. The parameter *scale* is the depth reconstruction's scale, which is described in subsection 3.2. Finally, the random pixel sample is merged with the depth discontinuity edge pixel to build the depth image's sparse representation. Here, the depth discontinuity edge is efficiently generated by the Laplace operator on the original depth image.

3.2. Depth Reconstruction

To recover the depth image, our method uses the sparse depth representation to reconstruct the depth image using an up- and down-sampling schema with Gaussian bilateral filtering. The up- and down-sampling procedure is based on the push-pull operation, which is described in [PHE*11].

The up-sampling procedure interpolates the high resolution image with the low resolution image using the upper scale's output to fill the current scale's holes, including the

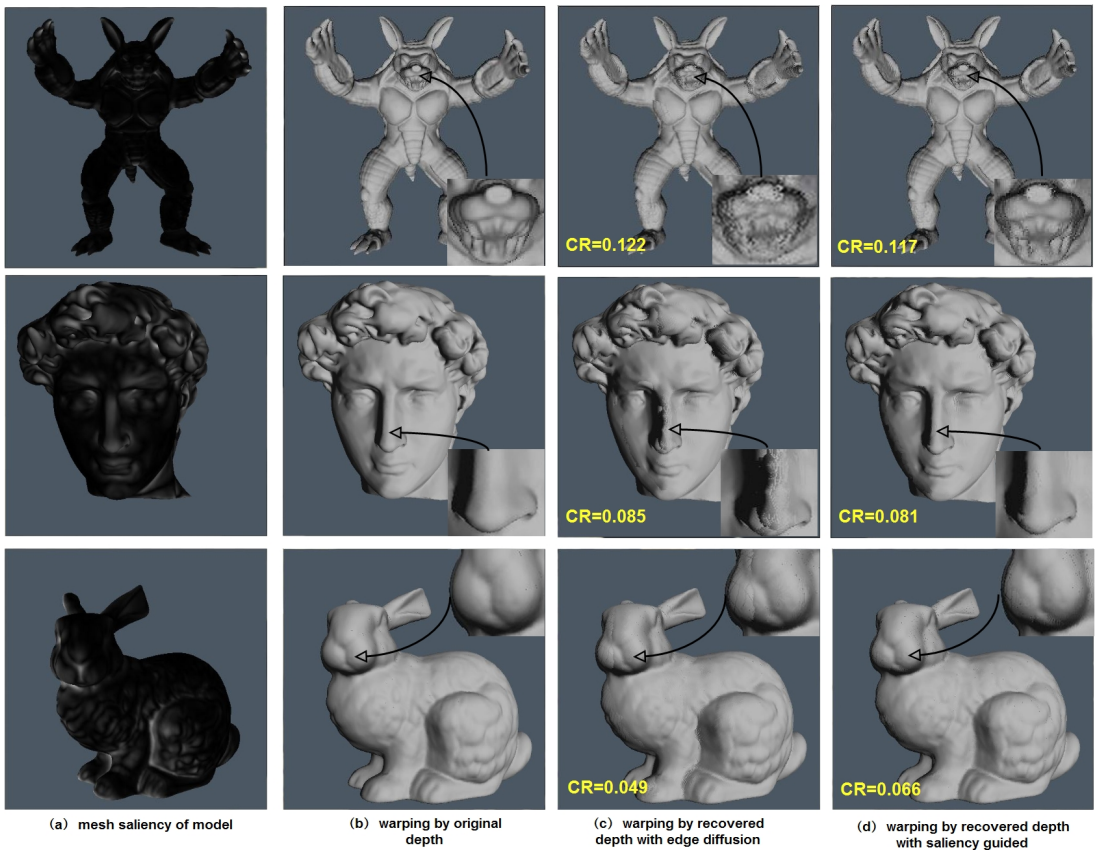

Figure 2: ***Multi-view 3D image warping****. In this case, we use depth images and color images from three referenced viewpoints to warp the current virtual viewpoint. The referenced viewpoint and virtual viewpoint are same for each group.*

diagonal diffusion and vertical/horizontal diffusion, and iterates until the scale is 0. However, the push-pull schema generates a jittery depth image in the smooth region of the recovered depth image, which can lead to a crack in the 3D image warping. Hence, before the coarse-scale depth images are pulled back to their corresponding position at the finer scale in the pull step, a bilateral filter smooths the upper scale while maintaining the precision of the depth discontinuity edge. The bilateral filter is formulated as

$$D(u) = \frac{1}{\sum W_{sr}} \sum_{j \in N_i} d_i \cdot f_s(i, j, scale) \cdot g_r(i, j) \qquad (2)$$

where N_i represents the i-th neighbor pixel domain, W_{sr} is the normalization factor, f_s is the spatial filter, which is defined as the Gaussian kernel and depends on the *scale*, and g_r is the domain filter

$$g_r(i, j) = (max(\varepsilon, 1 - \frac{|d_i - d_j|^2}{\sigma_z^2}))^r \qquad (3)$$

where σ_z is a normalized factor. In this paper, we set the value of σ_z as five percent of the view-frustum's length between the near and far planes. Our method defines five scales to reconstruct the depth image. In addition, the spatial filter f_s is dependent on the different scales. The method sets the Gaussian weighting marks $\sigma = 1$ and $N = 3$ when the scale is 3, 4 or 5 and $\sigma = 2$ and $N = 5$ when the scale is 0, 1 or 2, which can yield a visually pleasing result.

4. Experiments and Comparisons

Our work is performed on a Dell Precision T7600 with a 16-core Intel Xeon CPU E5-2687 Core(TM) 3.1 GHz and a NVIDIA Quadro 6000 in the OpenGL 4.0 environment. To achieve the effect on the sampling and reconstruction of the depth image, the method uses the RTT (render to texture) technique to obtain the depth image and the mesh's salient region to build the sparse depth representation. The recovered depth image is applied to the 3D image warping. Finally, the result of the 3D image warping is evaluated through visual perception and quantitative analysis.

4.1. Visual Evaluation

To visually test the quality of the recovered depth, the recovered depth image is used for 3D image warping. In this case, the final frame of the desired viewpoint is generated from three frames at several referenced viewpoints by multi-view 3D image warping. The visual impact of the desired viewpoint is explicitly influenced by the precision of the recovered depth. Moreover, to show the reliability and stability of the proposed method, we choose different models to generate depth images for sampling and reconstruction. To maintain the uniformity of the near and far plane of the frustum culling, the models are scaled into a bounding box with a length between 0 and 1. To facilitate a comparison to the edge diffusion method [PHE*11], the results are shown in Figure 2. The recovered depth image's warping results from our method are more similar to the original depth than the results of the edge diffusion method. The armadillo mesh's salient regions (head, hands and nose), the statue of David's three features (nose, left face and hair) and the bunny's important regions (mouth and leg) can be preserved better using the proposed method. To obtain the results of shown in Figure 2, our method's parameters are set to $r_1 = 0$ and $r_2 = 8$, and the threshold of the Laplace operator for detecting the depth discontinuity edge is $t = 0.0025$. In the edge diffusion method, the threshold t is set to 0.001. Here, the compression ratio is defined as

$$CR = \frac{D_{sparse}}{D_{original}}. \quad (4)$$

where $D_{original}$ is the number of valid pixels in the original depth image, and D_{sparse} is the number of valid pixels in the sparse depth representation.

4.2. Quantitative Evaluation

The result of the 3D image warping process can be evaluated quantitatively using the peak-signal-to-noise ratio (PSNR). The computation of PSNR uses the warping output from the original depth as the original image $I_{original}$ and the warping output from the recovered depth as the reconstructed image $I_{recovered}$. The results for the different models (armadillo, David and bunny) are shown in Figure 3. When the compression ratio is same for each model, the average PSNR is 0.2 dB higher than the depth edge diffusion method [PHE*11]. Nevertheless, our method cannot easily control the compression ratio, but it can adjust the sampling density to adapt to the approximate desired compression ratio. The proposed method has a low compression ratio for a single depth image; the average compression ratio of less than 15 percent for these models can achieve a high quality of synthetic images in 3D image warping.

4.3. Computation Complexity

To determine the efficiency of the proposed method, the runtime is tested for different resolutions (480*640, 600*800 and 768*1024). The experiment obtains the average time for 10 runs. As shown in Table 1, the sampling runtime approximates the efficiency of the sparse depth representation construction. The reconstruction shown in this table represents the depth reconstruction stage. The adaptive random sampling procedure requires additional computational cost; it relies on the efficiency of Poisson adaptive sampling, which requires the multi-pass rendering to generate the random pixel sample. Nevertheless, it reduces the number of pixel samples considerably while preserving the model's feature for 3D image warping. The efficiency of edge diffusion is superior to the proposed method. However, the purpose of the proposed method is different from that of the edge diffusion method. Our method mainly focuses on the quality of synthetic images for interactive 3D image warping, while the edge diffusion method focuses on rapidly compressing and decompressing depth images in the rendered video stream.

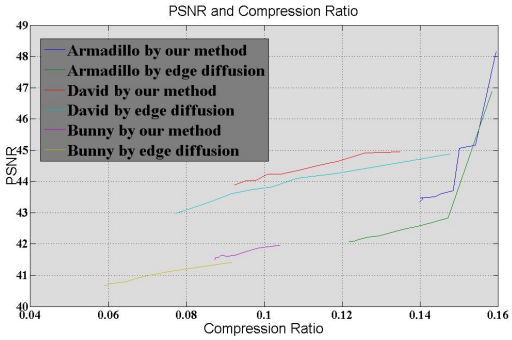

Figure 3: ***PSNR and compression ratio.*** *The PSNR represents the warping quality, and the compression ratio is defined as a ratio between the sampled depth and the original depth.*

Table 1: *Runtimes of different periods*

Resolution	Sampling	Reconstruction	Total
480*640	141 ms	5 ms	146 ms
600*800	291 ms	5 ms	206 ms
768*1024	311 ms	5 ms	316 ms

5. Conclusion and Future Work

This paper presented a new method of saliency-driven depth compression that combines mesh saliency and 2D features to guide the adaptive compression. In the the depth reconstruction stage, the proposed method exploits an up- and down-sampling schema with Gaussian bilateral filtering to recover the depth. The results showed that the method could preserve the precision in the mesh's salient region of the depth image and provide a beneficial visual impact for 3D image warping.

Limitations. The proposed method considered low-level human visual system cues to compute the mesh saliency to guide the depth compression. In addition, the proposed method is dependent on the mesh saliency to augment the 2D features to preserve the precision in the recovered depth, which is suitable for image-based rendering of virtual scenes.

Future work. In the future, it will be necessary to explore other information that can be used to detect important regions to preserve the higher level semantic information.

ACKNOWLEDGMENTS

This work is supported by the National High Technology Research and Development Program of China (No.2013AA013701). The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

References

[AMAM13] ANDERSSON M., MUNKBERG J., AKENINE-MÖLLER T.: Stochastic depth buffer compression using generalized plane encoding. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 103–112. 1

[Coo86] COOK R. L.: Stochastic sampling in computer graphics. *ACM Transactions on Graphics (TOG) 5*, 1 (1986), 51–72. 2

[CSS02] CHAI B.-B., SETHURAMAN S., SAWHNEY H. S.: A depth map representation for real-time transmission and view-based rendering of a dynamic 3d scene. In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on* (2002), IEEE, pp. 107–114. 1, 2

[FMY*13] FU J., MIAO D., YU W., WANG S., LU Y., LI S.: Kinect-like depth data compression. *Multimedia, IEEE Transactions on 15*, 6 (2013), 1340–1352. 1

[GGSC96] GORTLER S. J., GRZESZCZUK R., SZELISKI R., COHEN M. F.: The lumigraph. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), ACM, pp. 43–54. 2

[KCTS01] KRISHNAMURTHY R., CHAI B.-B., TAO H., SETHURAMAN S.: Compression and transmission of depth maps for image-based rendering. In *Image Processing, 2001. Proceedings. 2001 International Conference on* (2001), vol. 3, IEEE, pp. 828–831. 1

[KKS*05] KHAN A., KOMALO B., STAM J., FITZMAURICE G., KURTENBACH G.: Hovercam: interactive 3d navigation for proximal object inspection. In *Proceedings of the 2005 symposium on Interactive 3D graphics and games* (2005), ACM, pp. 73–80. 1

[KTL*04] KOLLER D., TURITZIN M., LEVOY M., TARINI M., CROCCIA G., CIGNONI P., SCOPIGNO R.: Protected interactive 3d graphics via remote rendering. In *ACM Transactions on Graphics (TOG)* (2004), vol. 23, ACM, pp. 695–703. 1, 2

[LD08] LAGAE A., DUTRÉ P.: A comparison of methods for generating poisson disk distributions. In *Computer Graphics Forum* (2008), vol. 27, Wiley Online Library, pp. 114–129. 2

[LKR*96] LINDSTROM P., KOLLER D., RIBARSKY W., HODGES L. F., FAUST N., TURNER G. A.: Real-time, continuous level of detail rendering of height fields. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), ACM, pp. 109–118. 1, 2

[LVJ05] LEE C. H., VARSHNEY A., JACOBS D. W.: Mesh saliency. In *ACM Transactions on Graphics (TOG)* (2005), vol. 24, ACM, pp. 659–666. 2

[Mar02] MARCELLIN M. W.: *JPEG2000: image compression fundamentals, standards, and practice*, vol. 1. springer, 2002. 1, 2

[MF*06] MORVAN Y., FARIN D., ET AL.: Platelet-based coding of depth maps for the transmission of multiview images. In *Electronic Imaging 2006* (2006), International Society for Optics and Photonics, pp. 60550K–60550K. 1, 2

[MFY*09] MORI Y., FUKUSHIMA N., YENDO T., FUJII T., TANIMOTO M.: View generation with 3d warping using depth information for ftv. *Signal Processing: Image Communication 24*, 1 (2009), 65–72. 1

[MJ97] MCMILLAN JR L.: *An image-based approach to three-dimensional computer graphics*. PhD thesis, Citeseer, 1997. 2

[MMB97] MARK W. R., MCMILLAN L., BISHOP G.: Post-rendering 3d warping. In *Proceedings of the 1997 symposium on Interactive 3D graphics* (1997), ACM, pp. 7–ff. 2

[NMD13] NGUYEN V.-A., MIN D., DO M. N.: Efficient techniques for depth video compression using weighted mode filtering. *Circuits and Systems for Video Technology, IEEE Transactions on 23*, 2 (2013), 189–202. 1

[PHE*11] PAJAK D., HERZOG R., EISEMANN E., MYSZKOWSKI K., SEIDEL H.-P.: Scalable remote rendering with depth and motion-flow augmented streaming. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 415–424. 1, 2, 3, 5

[PHM*14] PAJĄK D., HERZOG R., MANTIUK R., DIDYK P., EISEMANN E., MYSZKOWSKI K., PULLI K.: Perceptual depth compression for stereo applications. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 195–204. 1

[PJO*09] PARK Y. K., JUNG K., OH Y., LEE S., KIM J. K., LEE G., LEE H., YUN K., HUR N., KIM J.: Depth-image-based rendering for 3dtv service over t-dmb. *Signal processing: Image communication 24*, 1 (2009), 122–136. 1

[QLLM13] QUINN J. A., LANGBEIN F. C., LAI Y.-K., MARTIN R. R.: Generalized anisotropic stratified surface sampling. *Visualization and Computer Graphics, IEEE Transactions on 19*, 7 (2013), 1143–1157. 2

[SGHS98] SHADE J., GORTLER S., HE L.-W., SZELISKI R.: Layered depth images. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), ACM, pp. 231–242. 2

[Smo11] SMOLIC A.: 3d video and free viewpoint video-from capture to display. *Pattern recognition 44*, 9 (2011), 1958–1968. 1

[SMW07] SCHWARZ H., MARPE D., WIEGAND T.: Overview of the scalable video coding extension of the h. 264/avc standard. *Circuits and Systems for Video Technology, IEEE Transactions on 17*, 9 (2007), 1103–1120. 1

[TW08] TZENG S., WEI L.-Y.: Parallel white noise generation on a gpu via cryptographic hash. In *Proceedings of the 2008 symposium on Interactive 3D graphics and games* (2008), ACM, pp. 79–87. 3

[Wei08] WEI L.-Y.: Parallel poisson disk sampling. In *ACM Transactions on Graphics (TOG)* (2008), vol. 27, ACM, p. 20. 2, 3

[WYT*10] WILDEBOER M. O., YENDO T., TEHRANI M. P., FUJII T., TANIMOTO M.: Color based depth up-sampling for depth compression. In *Picture Coding Symposium (PCS), 2010* (2010), IEEE, pp. 170–173. 1, 2