

Objective quality analysis for free-viewpoint DIBR

Citation for published version (APA):

Do, Q. L., Zinger, S., & With, de, P. H. N. (2010). Objective quality analysis for free-viewpoint DIBR. In *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP 2010), 26-29 September 2010, Hong Kong, Hong Kong* (pp. 2629-2632). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/ICIP.2010.5652546>

DOI:

[10.1109/ICIP.2010.5652546](https://doi.org/10.1109/ICIP.2010.5652546)

Document status and date:

Published: 01/01/2010

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

OBJECTIVE QUALITY ANALYSIS FOR FREE-VIEWPOINT DIBR

Luat Do^a, Svitlana Zinger^a and Peter H.N. de With^{a,b}

^aEindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, Netherlands

^bCyclomedia Technology B.V., P.O. Box 68, 4180 BB Waardenburg, The Netherlands

{Q.L.Do, S.Zinger, P.H.N.de.With}@tue.nl

ABSTRACT

Interactive free-viewpoint selection applied to a 3D multi-view video signal is an attractive feature of the rapidly developing 3DTV media. In recent years, significant research has been done on free-viewpoint rendering algorithms which mostly have similar building blocks. In this paper, we analyze the four principal building blocks of most recent rendering algorithms and their contribution to the overall rendering quality. We have discovered that the rendering quality is dominated by the first step, *Warping* which determines the basic quality level of the complete rendering chain. The third step, *Blending*, is a valuable step which further increases the rendering quality by as much as 1.4 dB while reducing the disocclusions to less than 1% of the total image. Varying the angle between the two reference cameras, we notice that the quality of each principal building block degrades with a similar rate, 0.1–0.3dB/degree for real-life sequences. While experimenting with synthetic data of higher accuracy, we conclude that for developing better free-viewpoint algorithms, it is necessary to generate depth maps with more quantization levels so that the *Warping* and *Blending* steps can further contribute to the quality enhancement.

Index Terms— Free-viewpoint, DIBR, 3DTV, PSNR analysis

1. INTRODUCTION

Three-dimensional television (3DTV) is nowadays broadly considered to succeed existing 2D HDTV technology. An interesting future feature in 3D imaging is to virtually move through the scene in order to create different viewpoints. This feature, called free-viewpoint rendering in multi-view video has become a popular topic in 3D research that can lead to applications such as free-viewpoint TV (FTV) [1], 3D medical imaging [2], multimedia services [3] and 3D reconstruction [4]. Since the number of cameras is practically limited and thereby the number of viewing angles, research has been devoted to interpolate views between the cameras (view synthesis). A well-known technique for view synthesis is called Depth Image Based Rendering (DIBR), which involves the projection, or warping, of one viewpoint into another driven by a dense depth map. Most recent algorithms [5, 6, 7, 8, 9, 10] employ projection from two reference views to generate a virtual one. Although small differences exist, these algorithms consist of basically the same building blocks. In the first step, textures and depth values are projected to the virtual view. The second step closes holes and processes contour artifacts that projection produces. The third step blends the two projected views and in the last step the remaining disocclusions are inpainted. At present, the rendering quality is measured by analyzing the output of the algorithm. Few quality assessments are performed on the individual blocks. In this paper, we perform a comprehensive

and detailed analysis of the individual building blocks of most recent free-viewpoint DIBR algorithms. Furthermore, we analyze the contribution of each block to the overall rendering quality, since it provides insight about the most critical processing steps and leads to ways for improvement. For measuring the rendering quality of each building block, we employ the PSNR metric. Other objective quality metrics exist, such as described in [11, 12, 13]. These methods try to correlate the measurements with the perceived quality. Nevertheless, we have adopted the PSNR as the quality metric for comparison for two reasons. First, an advantage of the PSNR metric is that errors are measured on pixel basis so that errors resulting from the projection and interpolation will also contribute to the measured PSNR. Second, it is the most proven and commonly used method to measure quality differences in sequences with distributed moderate degradation. It will be shown that our analysis reveals that the quality improvement for DIBR concentrates on basically two processing steps that depend on the accuracy of the depth signal.

The remainder of this paper is organized as follows. Section 2 gives an overview of a representative free-viewpoint DIBR algorithm used for experimenting. Section 3 explains the experimental quality analysis for free-viewpoint DIBR and the applied test sequences. Section 4 shows the obtained results and finally, Section 5 discusses the conclusions and recommendations for next generation free-viewpoint DIBR algorithms.

2. FREE-VIEWPOINT DIBR ALGORITHM

Let us briefly describe our own recent DIBR algorithm [10] and analyze the quality of each processing step. This algorithm contains similar processing steps as other recent DIBR algorithms [5, 6, 7, 8, 9], and we have the detailed description available. It is a simple and fast algorithm that yields results comparable to most recent algorithms. Our algorithm, depicted in Fig. 1, consists of four steps and the individual steps will now be explained. For a more detailed description of this algorithm, we refer to our paper [10].

Step 1, *Warping*: In the first step, we project or warp the textures and depth maps at the reference views to the virtual position. Contour artifacts are treated by finding high discontinuities prior to warping and omit those pixels from projecting.

Step 2, *Cracks*: Warping causes holes, or cracks, to appear at the virtual image. In this step, these holes are labeled by a median filter. The texture and depth values for the cracks are then retrieved by inverse warping.

Step 3, *Blending*: In this step, the two projected images are blended to ensure that foreground objects are visible. Blending also decreases errors in viewpoint dependency of textures and reduces disocclusions.

Step 4, *Inpainting*: In the last step, the remaining black pixels

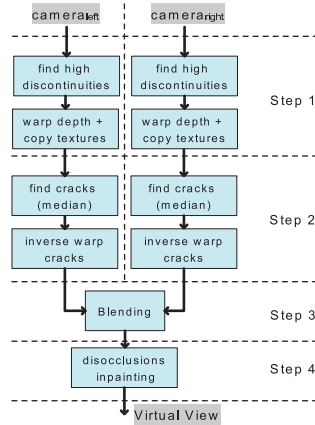


Fig. 1: Sequence of the principal steps in our rendering algorithm.

are inpainted by a depth-assisted weighted interpolation of textures from the edges of disocclusions.

3. SETUP FOR QUALITY ANALYSIS

In this section, we describe our quality analysis approach for the individual steps of the algorithm, the conducted multiple experiments and the five chosen test sequences.

For creating a ground truth, we generate a virtual view on a position corresponding with one of the reference cameras, using the two nearest camera views for projection. The PSNR of the luminance can then be calculated from the mean squared error between the virtual and the ground-truth view. The rendering quality is calculated as an average over all 100 frames of the sequence. For each individual step of the algorithm, we measure the PSNR for only those pixels to which the step is applied.

In the first series of experiments we measure the rendering quality of each individual step of the algorithm for those pixels influenced by that individual step. The performance of each step and possible bottlenecks in the algorithm can then be analyzed. In the second series of experiments, the accumulated rendering quality after each step is calculated. The influence of each individual step on the overall rendering quality can then be analyzed and visualized. It should be noted that the angle between two reference cameras (γ) is an important factor in 3D multi-view video. Therefore, we have performed the above two series of experiments for a varying γ , as described in [8].

The test sequences are depicted in Fig 2. The two well-known real-life sequences, ‘Ballet’ and ‘Breakdancers’, are chosen for reference purposes. The depth maps of these two sequences contain errors and therefore we have created three synthetic sequences (‘Cubes’, ‘Cubes_low’ and ‘Room’) to experiment with accurate depth maps. The ‘Cubes’ sequence consists of three cubes rotating and circling in the foreground in a room mapped with high-frequency textures. To measure the influence of high-frequency textures on the view synthesis, we have created the same ‘Cubes’ sequence, but now the room is mapped with low-frequency textures. The ‘Room’ sequence is created to measure the influence of foreground objects on the view synthesis. This sequence is the same as the ‘Cubes’ sequence, but the cubes are removed, so there are no foreground objects in this sequence. All sequences consist of 100 frames with 8 different camera views, spanning about 30° degrees from one end to the other. The image size is 1024×768 pixels with 8-bit precision texture and depth values.

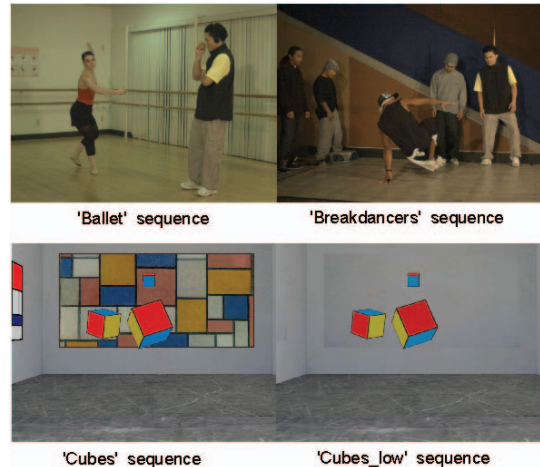


Fig. 2: Real-life (above) and synthetic (below) test sequences.

4. EXPERIMENTAL RESULTS

Table 1 contains the results for view synthesis with Camera 3 as the virtual viewpoint and Cameras 2 and 4 as the left and right reference cameras, respectively. The second and third column contain the obtained rendering quality of each individual step for the texture and depth maps, respectively. The fourth column denotes the amount of pixels involved with that processing step. The fifth column contains the accumulated rendering quality of the algorithm after each step for texture maps and the last column denotes the accumulated amount of pixels that is covered up to and including that processing step. The last two columns of the ‘Ballet’ sequence are visualized in Fig. 3 and provide an interesting view on the quality development.

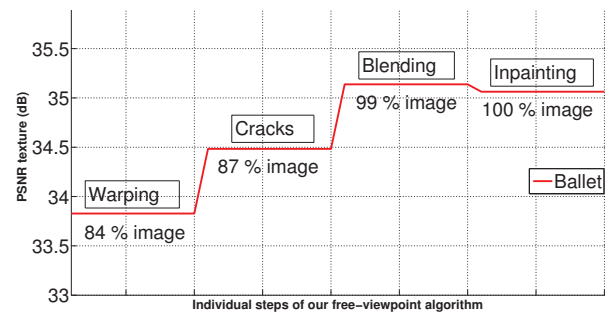


Fig. 3: Quality development of texture after each individual step of our free-viewpoint algorithm for the ‘Ballet’ sequence.

Let us now discuss the obtained results. Fig. 3 shows that the resulting rendering quality is largely determined by the first step of the algorithm. This is not surprising since this step, *Warping*, operates on 80–90% of the pixels. Furthermore, it is clear that blending also gives a valuable contribution. This step not only increases the rendering quality by as much as 1.4 dB, but it also decreases the disocclusions from roughly 10% to less than 1% of the image. Focussing on Step 2 (*Cracks*), we see that the quality is lower than that of Step 1 because at Step 2, we perform inverse warping driven by only those depth values that are estimated by a median. Nevertheless, we notice that the rendering quality after Step 2 is higher than after Step 1 for both real-life sequences. This can be explained by the correcting behaviour of the median filter. This filter not only closes holes

Table 1: Breakdown of the individual steps of our free-viewpoint DIBR algorithm; quality and number of pixels.

Ballet	PSNR (texture)	PSNR (depth)	% pixels	PSNR,step (texture)	% pixels
Warping	33.8	30.4	84.2	33.8	84.2
Cracks	32.2	31.0	5.0	34.5	87.3
Blending	35.1	30.7	99.1	35.1	99.1
Inpainting	30.7	25.8	0.9	35.1	100
Breakdancers	PSNR (texture)	PSNR (depth)	% pixels	PSNR,step (texture)	% pixels
Warping	33.9	37.1	90.6	33.9	90.6
Cracks	32.1	35.8	5.8	33.9	94.3
Blending	35.3	36.4	99.3	35.3	99.3
Inpainting	23.7	27.7	0.7	34.9	100
Cubes (high texture)	PSNR (texture)	PSNR (depth)	% pixels	PSNR,step (texture)	% pixels
Warping	32.4	48.8	89.9	32.4	89.8
Cracks	26.0	45.5	2.5	32.1	91.6
Blending	32.5	48.4	99.3	32.5	99.3
Inpainting	27.6	31.8	0.7	32.5	100
Cubes_low (low texture)	PSNR (texture)	PSNR (depth)	% pixels	PSNR,step (texture)	% pixels
Warping	35.4	48.4	89.8	35.4	89.8
Cracks	33.7	45.5	2.5	35.3	91.6
Blending	36.0	48.4	99.3	36.0	99.3
Inpainting	30.9	31.8	0.7	35.9	100
Room	PSNR (texture)	PSNR (depth)	% pixels	PSNR,step (texture)	% pixels
Warping	32.9	58.3	90.8	32.9	90.8
Cracks	26.1	53.4	2.4	32.6	92.5
Blending	33.0	57.8	99.4	33.0	99.4
Inpainting	28.4	32.0	0.6	33.0	100

but also corrects pixels that are warped with a wrong depth value, which occurs for the ‘Ballet’ and ‘Breakdancers’ sequences. For the synthetic sequences, the rendering quality after Step 2 is worse than after Step 1, which is expected, because the depth maps of these sequences contain no errors. The performance contribution of the inpainting step to the overall PSNR is the lowest, but since the disoccluded regions are less than 1% of the image, the contribution of this step to the overall rendering quality is very small. However, the subjective quality is significantly improved as unfilled holes give an annoying artifact.

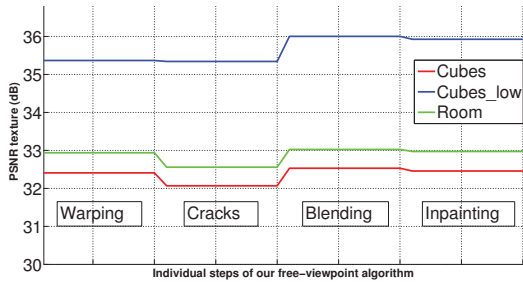


Fig. 4: Quality development of texture after each individual step of our free-viewpoint algorithm for the synthetic sequences.

Next, we investigate the influence of high-frequency textures and foreground objects on the rendering quality by comparing the results of the synthetic sequences. In Fig. 4 and 5, the rendering quality after each step is visualized for texture and depth maps, respectively. Comparing the ‘Room’ sequence with the two cubes sequences, we observe that the rendering quality for depth maps is highly dependent

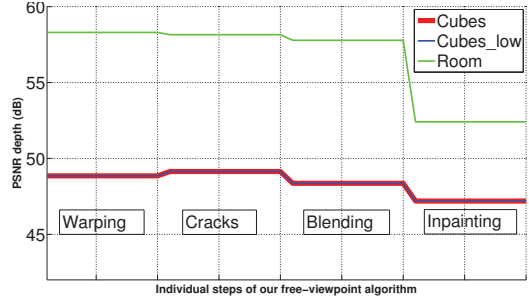


Fig. 5: Quality development of depth after each individual step of our free-viewpoint algorithm for the synthetic sequences.

on foreground objects. Even though the sizes of the three cubes are relatively small, the errors made at the edges of these objects with projecting cause a quality drop of more than 5 dB. By comparing the two cubes sequences with low and high-frequency textures, we notice that the choice of high-frequency textures alone causes a quality drop of 3 dB. The ‘Room’ sequence experiences a quality increase of 0.5 dB over the ‘Cubes’ sequence because the sequence does not contain any foreground objects. From Fig. 4, we conclude that the rendering quality of textures at the virtual image is dependent on the amount of foreground objects but even more on the complexity of occurring textures in the scene. Even though the ‘Cubes’ sequence contains no depth map errors, the rendering quality for textures is not higher than for the ‘Ballet’ sequence. This leads us to the conclusion that the projection, or warping, produces many errors due to the discrete quantization levels of depth maps. Our assumptions are visually supported when observing the error maps of the virtual views as shown in Fig. 6. For the texture values, the errors are located at edges of foreground objects and at areas of high-frequency textures, while the depth map errors are only located around edges of foreground objects.

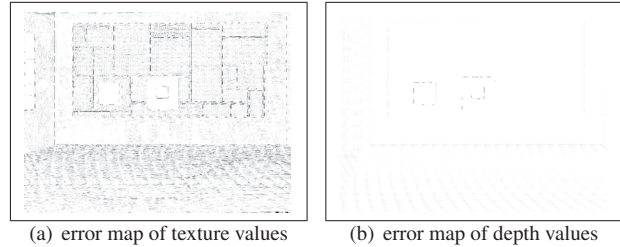


Fig. 6: Error maps for texture and depth values of the ‘cubes’ sequence (white no error, black error).

Let us now discuss the influence of the angle between the two reference cameras. The results for the ‘Ballet’ sequence and the ‘Cubes’ sequence are depicted in Fig. 7 and Fig. 8, respectively. For an increasing angle, we can observe that the quality produced by the individual steps degrades with a similar rate. For the ‘Ballet’ and ‘Cubes’ sequence, the degradation is 0.3 dB/degree and 0.05 dB/degree, respectively. The degradation for the ‘Ballet’ sequence is larger because the sequence contains much larger foreground objects and the depth maps contain errors. Although the degradation per degree is small, generating virtual images with reference angles larger than 15° are not recommended because of the large disoccluded areas. At these large angles, the disocclusions consist of more than 2% of the image and cause annoying artifacts, such as low-frequency patches and color degradation.

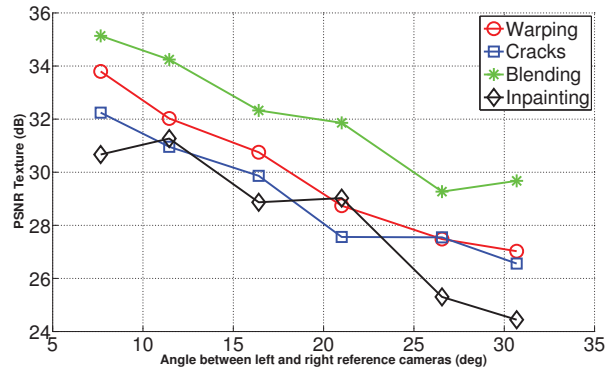


Fig. 7: Texture quality of each individual step as a function of the angle between two reference cameras for the ‘Ballet’ sequence.

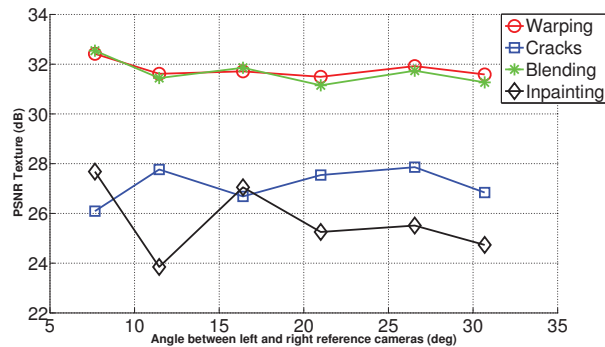


Fig. 8: Texture quality of each individual step as a function of the angle between two reference cameras for the ‘Cubes’ sequence.

5. CONCLUSIONS

This paper has presented a performance analysis of the steps in a free-viewpoint DIBR algorithm for rendering virtual views in a 3D multi-view video sequence. The algorithm is representative as it contains the commonly used processing steps presented in literature, such as warping, cracks, blending and inpainting. Experiments were conducted for both natural and synthetic sequences and concentrated on the following aspects: (1) The contribution of each individual step of the algorithm by measuring the rendering quality and the amount of pixels to which this step is applied, (2) the contribution of each step of the algorithm to the overall rendering quality, (3) the angle dependency between two reference cameras. For completeness, the experiments were performed on both texture and depth maps of the virtual view. From the obtained results, we have concluded the following for the individual steps of the algorithm. First, the starting step, which is *Warping*, determines largely the overall rendering quality, since this step operates on 80–90% of the virtual image. Second, the *Blending* step can increase the quality by as much as 1.4 dB and reduce the disocclusions from 10% to less than 1% for small angles. Third, the *Cracks* and *Inpainting* steps slightly decrease the PSNR quality, but these steps operate on such a small amount of pixels that the overall quality is not affected significantly. However, these steps give a notable subjective quality improvement. Fourth, the performance of the individual algorithm steps degrades with a similar rate when the angle between reference cameras increases. For the real-life and synthetic sequences, the degradation is 0.1–0.3 dB/degree and 0.05 dB/degree, respectively. In addition, experiments with the three synthetic sequences, which differ in texture

quality and foreground object complexity, provided new insights in the projection and interpolation errors on two aspects. First, the errors created during (*Warping*) are caused by the limited quantization levels in depth maps. Second, by inspecting the errors at the virtual texture and depth maps, we learn that the rendering quality for depth maps is limited only by the amount of foreground object edges as errors are concentrated on the edges. On the other hand, the quality of texture images is limited not only by the amount of foreground objects but also by the quality of textures in the reference views.

Our detailed quality analysis gives clues for developing an improved class of next generation view synthesis rendering algorithms. First of all, to obtain a higher rendering quality, the projection errors in the first step need to be minimized. This is possible by generating depth maps with more quantization levels, either on the capturing side, or at the rendering side. Another gain in rendering quality can be expected by specifically processing edges of foreground objects. For example, algorithms may adapt their sampling rate by performing sub-pixel projection to the virtual viewpoint at these edges to achieve a higher precision and thus higher quality.

6. REFERENCES

- [1] M. Tanimoto, “FTV (free viewpoint television) for 3D scene reproduction and creation,” in *CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, Washington, DC, USA, 2006, p. 172, IEEE Computer Society.
- [2] D. Ruijters and S. Zinger, “IGLANCE: transmission to medical high definition autostereoscopic displays,” in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2009.
- [3] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, “Multiview imaging and 3DTV,” *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 10–21, 2007.
- [4] C. Leung and B. C. Lovell, “3D reconstruction through segmentation of multi-view image sequences,” *Workshop on Digital Image Computing*, vol. 1, pp. 87–92, 2003.
- [5] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, “High-quality video view interpolation using a layered representation,” in *ACM SIGGRAPH 2004 Papers*, New York, NY, USA, 2004, pp. 600–608, ACM.
- [6] C. Fehn, “Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV,” in *Stereoscopic Displays and Virtual Reality Systems XI*, May 2004, vol. 5291, pp. 93–104.
- [7] A. Smolic, K. Müller, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, “Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems,” in *ICIP*. 2008, pp. 2448–2451, IEEE.
- [8] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, “View generation with 3D warping using depth information for FTV,” *Image Commun.*, vol. 24, no. 1-2, pp. 65–72, 2009.
- [9] K.-J. Oh, Y. Sehoon, and Y.-S. Ho, “Hole-filling method using depth based in-painting for view synthesis in free viewpoint television (ftv) and 3D video,” in *Picture Coding Symposium*, Chicago, USA, 2009.
- [10] L. Do, S. Zinger, and P.H.N. de With, “Quality improving techniques for free-viewpoint DIBR,” in *Stereoscopic displays and applications XXII*, 2010.
- [11] Zhou Wang, Ligang Lu, and Alan C. Bovik, “Video quality assessment based on structural distortion measurement,” *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121 – 132, 2004.
- [12] J. Starck, J. Kilner, J. Y. Guillemaut, and A. Hilton, “Objective quality assessment in free-viewpoint video production,” *Image Commun.*, vol. 24, no. 1-2, pp. 3–16, 2009.
- [13] H. Shao, X. Cao, and G. Er, “Objective quality assessment of depth image based rendering in 3DTV,” in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2009.