Suryanarayana M. Muddala; Mårten Sjöström; and Roger Olsson. "Edge-preserving depth-image-based rendering method ", in *International Conference on* 3D *Imaging 2012 (IC3D)*, 3-5 Dec. 2012.

DOI: 10.1109/IC3D.2012.6615113

EDGE-PRESERVING DEPTH-IMAGE-BASED RENDERING METHOD

Suryanarayana M. Muddala, Mårten Sjöström, Roger Olsson

Dept. of Information Technology and Media, Mid Sweden University, SE-85170 Sundsvall, Sweden

ABSTRACT

Distribution of future 3DTV is likely to use supplementary depth information to a video sequence. New virtual views may then be rendered in order to adjust to different 3D displays. All depth-imaged-based rendering (DIBR) methods suffer from artifacts in the resulting images, which are corrected by different post-processing. The proposed method is based on fundamental principles of 3D-warping. The novelty lies in how the virtual view sample values are obtained from one-dimensional interpolation, where edges are preserved by introducing specific edge-pixels with information about both foreground and background data. This avoids fully the post-processing of filling cracks and holes. We compared rendered virtual views of our method and of the View Synthesis Reference Software (VSRS) and analyzed the results based on typical artifacts. The proposed method obtained better quality for photographic images and similar quality for synthetic images.

Index Terms— 3D, video plus depth, warping, depth-image-based view rendering.

1. INTRODUCTION

Distribution of future three-dimensional (3D) TV may be based on the video plus depth (V+D) format, which allows to adjusting 3D contents to different 3DTV sets [1]. New virtual views may then be rendered to produce different views of the captured scene to create a multiscopic 3D-presentation. Common 3D rendering algorithms produce a number of artifacts, which are corrected with dedicated post-processing, and so the question rises if there exists a more fundamental approach that avoids these artifacts?

3DTV has become increasingly popular. Today it requires special glasses in order to separate left and right views, but 3DTV based on multiview displays allows experiencing the stereoscopic effect without spectacles as the displays distribute several perspective views into the viewing zone. An efficient and generic way to distribute 3D contents to end-users is by supplementing the 2D video by a depth value for each pixel. This V+D format enables rendering of new virtual views of the scene for these multiview displays [1]. An extension is to have multiple views with corresponding depth information (MVD), which permits virtual views to be rendered from two adjacent V+D sequences so that occluded areas from one sequence can be filled by the other sequence. Typical for distribution of 3DTV using these formats is that the sequences are internally aligned. Furthermore, the end devices are now only capable of presenting horizontal parallax, which leads to a synthesis of new virtual views along the horizontal axis only. The present work limits itself to rendering from the V+D format.

Depth-image-based rendering (DIBR) is based on perspective 3D warping [1,2]. The virtual views often contain artifacts such as empty cracks, translucent cracks, and empty regions, which are due to shortcomings of the DIBR method. Several solutions have been proposed in the literature to reduce the effect of these artifacts. Zitnick et al. divided the original data into layers to separate data depending on reliability [3], which was further improved in [4]. Zhang et al. applied an asymmetric pre-filter to the depth map [5]. Several improvements have lead to the view synthesis reference software (VSRS) [6], which is one of the best performing methods today. VSRS also includes a 1D fast mode to render horizontally displaced virtual views only.

The present work proposes a rendering method that relies on fundamental forward 3D warping and interpolation. The novelty of the method is the introduction of edge-pixels to keep track of both background and foreground information at edges, whereby interpolation is straightforward. The method further applies edge softening by low-pass filtering. Possible artifacts of 3D warping and their remedies were analyzed for both the proposed method and for the VSRS. The proposed method has a straightforward approach, yet it gives better quality than the VSRS for both computer generated images and photographic images.

The paper is outlined as follows. Section 2 explains the fundamentals of 3D warping and describes the problem. Section 3 presents the proposed method. The experimental set-up and the evaluation criteria are described in Section 4. The results and analysis follows in Section 5 and 6, and the conclusions are found in Section 7.

2. PROBLEM DESCRIPTION

The DIBR methods based on perspective 3D warping generate virtual views with a correct geometric perception. The V+D information and the camera parameters are then required to produce a virtual view [1, 2]. The extrinsic parameters express the necessary rotation $\mathbf{R}$ and translation $\mathbf{t}$ to transform a 3D point with homogeneous coordinates $\mathbf{M} = (x, y, z, 1)^T$ in the scene into a camera pixel $\boldsymbol{m} = (u/z, v/z, 1)^T$. The intrinsic parameters are represented by the calibration matrix $\mathbf{K}$ that describes the focal distance, image centre and camera pixels sizes. Together they define the projection matrix $\mathbf{P} = \mathbf{KI}[\mathbf{R}|\mathbf{t}]$. The forward warping produces a new pixel in the virtual view.

978-1-4799-1580-4/12/$31.00 ©2012 IEEE

$$z_v \mathbf{m}_v = z_o \mathbf{P}_v \mathbf{P}_o^{-1}(\mathbf{m_o}) \tag{1}$$

where, subscripts v and o corresponds to virtual and original views, respectively. A detailed description is found in [1, 2].

The 3D warping produces a number of artifacts in the virtual view; typical artifacts are listed in Table I along with their causes. Previous research works have proposed remedies for each individual artifact, giving a palette of tools to correct the errors. The aim of this work was instead to search for a DIBR method that relies on fundamental principles, which reduces its inherent artifacts, so that specific solutions are not required for each separate artifact.

Specifically, the objectives of the study were to identify artifacts due to DIBR by forward 3D warping and its remedies of previous algorithms; and to propose a simple DIBR method that has at least as good quality as the VSRS. We limited this first study to V+D format data only and focused it on rendering errors that have been identified in earlier works. We assumed virtual views to be restricted to a horizontal displacement with respect to the original view as this is applicable for 3DTV distribution. We also considered both computer generated and photographic contents in order to investigate its applicability to different 3D material.

TABLE I

Artifact	Cause	Remedy	
		VSRS	Proposed Method
Empty cracks	Integer round-offs of projected coordinates	Median filtering	Interpolation
Translucent cracks	Background pixels seep through cracks	Constraints on pixel mapping order	Removal of hidden pixel
Empty regions	Disocclusions	Background propagation	Introduction of edge-pixels, interpolation
Smearing of edges	Smooth texture edges or depth-texture misalignment	Removal of unreliable pixels before warping	Selection of "pure" background and foreground colors for edge-pixels
Unnatural contours	Pixilation at edges due to new border between foreground and background	Splatting along edges	Low pass filtering along edges

3. PROPOSED METHOD

The method we propose to solve the stated problem is based on forward 3D warping. The pixels projected into the virtual view keep the exact position in a first stage, and are not assigned to the closest virtual pixel as in common DIBR algorithms. A subsequent interpolation of projected samples gives the values of the pixel grid in the virtual view. Specific edge-pixels are introduced at borders between foreground and background. They are assigned foreground depth but contain both background and foreground colors at the corresponding side of the edge. These edge-pixels keep track of the placement of edges and the color of the respective side, and so avoid smearing out edges over disocclusions in the interpolation process. We further apply edge softening by a low-pass filtering of edges.

The assumption of only horizontal displacement reduces the previously described warping to a one-dimensional translation. The proposed method thereby turns into a line-by-line algorithm with the following steps depicted in Fig. 1:

Edge detection: The difference between neighboring pixels on the depth map is calculated, which identifies edges between foreground and background or objects at intermediate levels. Because the detection is made along one line, the detection narrows down to finding a difference larger than a threshold.

Edge pixels: Introduces an edge-pixel containing foreground depth and both background and foreground colors at the corresponding side of the pixel. See Fig. 2. This step ensures correct information on each side of an edge. The color must be of complete background and foreground colors, respectively, in order to avoid smearing colors over a larger area in the warping process. Therefore, color information is collected 1-2 pixels from each side of the edge. The edge-pixel is practically implemented by adding a pixel horizontally shifted by a small amount ε .

3D Warping: Applies geometrical forward warping to find the new floating point coordinates for each pixel in the virtual view.

Hidden pixels removal: Deletes the pixels that become occluded in the new virtual view. Hidden pixels are identified by subtracting the consecutive pixels' virtual view x-coordinates: a negative result indicates a hidden pixel.

Interpolation: Assigns values to the pixel grid of the virtual view by applying arbitrary interpolation method to the rendered pixels with floating point coordinates.

Edge Smoothing: Applies a low pass filter of color values over a small area around all edges. The edges' positions are known by re-using positions from the previous edge detection and the subsequent warping. This smoothing counteracts the pixilation that occurs at the edges.

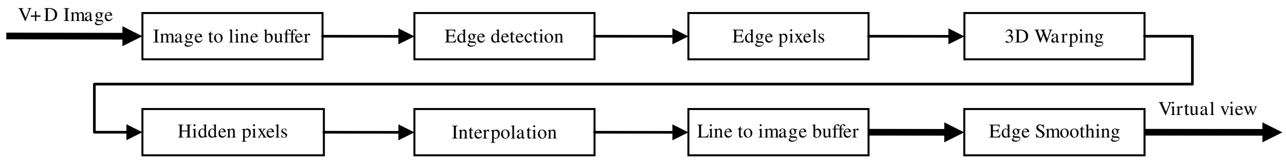


Fig. 1. Processing steps of the proposed method. Horizontal displacement only allows for line-by-line processing, with the exception of the final edge smoothing step.

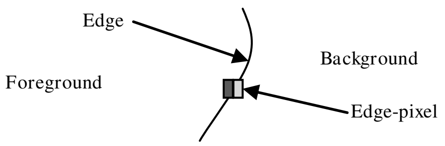


Fig. 2. Edge-pixels are introduced at borders between foreground and background. They contain only foreground depth but both background and foreground colors at the corresponding side of the pixel.

4. TEST ARRANGEMENT AND EVALUATION CRITERIA

The proposed method was assessed by computing virtual views and comparing those to captured views at the same position. The displacement of the virtual camera was parallel to the given view by an amount of one ocular distance (65 mm). Equation (1) can then be simplified to

$$u_v = u_o + f \cdot \left(t_{x,v} - t_{x,o}\right) / z_o + \left(o_{x,v} - o_{x,o}\right), \qquad (2)$$

where f is the focal length; $t_{x,v}$ and $t_{x,o}$ are horizontal component of the translation vector t for virtual and original views respectively; $o_{x,v}$ and $o_{x,o}$ are the principal component offset for the virtual and original views. The proposed method was implemented in Matlab.

Edge-pixels were implemented by adding an extra pixel with horizontal shift $\varepsilon = \pm 0.01$, where the sign is given by the depth map derivative at the edge. We applied both nearest neighbor and linear interpolations. The low pass edge smoothing filter was a combination of bicubic interpolation and an averaging filter.

The view quality achieved by the proposed method was compared to the quality obtained by VSRS using its 1D approach, i.e. line-by-line rendering. VSRS have incorporated many tools proposed in the literature to correct rendering artifact (see Table I). In this way, VSRS can be considered state-of-the-art of DIBR methods. As VSRS is developed for the MVD format, a virtual view was rendered from a single V+D sequence by providing the software with the same sequences and the same camera settings for both left and right view input.

The input data used for the assessment were both computer generated and photographic. The former contains exact information about edges and depths by default, and further consisted of smooth surfaces. The latter consisted of an extract of the first 10 frames of the MVD sequences "Lovebird" and "Poznan Street" [7]. "Lovebird" has larger depth discontinuities and "Poznan Street" has a complex depth distribution and also a large amount of structural variations in the texture. The Lovebird sequence's view 6 was used to compute virtual view at the position of view 4 and Poznan street sequence's view 4 was used to compute virtual views at the position of view 3 and 5, respectively, in order to have a ground truth for comparison.

The rendered views were compared to the corresponding originals using Peak Signal-to-Noise-Ratio for the intensity component (YPSNR) and Mean Structural Similarity Index (MSSIM) [8]. Both metrics measure the view quality, where the latter better corresponds to the experienced visual quality. The visual quality was also examined by direct inspection of the views. The execution times of the methods were not considered because the methods were implemented in different platforms. The proposed method was tested on a frame-by-frame basis not considering time effects.

5. RESULTS

The outcome of using photographic input sequences is shown in Fig. 3. The YPSNR and MSSIM graphs consistently demonstrate that the proposed method with linear interpolations performs better than with nearest neighbor interpolation and better than the VSRS In one case, nearest neighbour interpolation resulted in a similar quality to the result of the VSRS, see Fig. 3(c). Fig. 4 (a) and (b) depict details of a rendered frame at view position 3. It exhibits the visual improvements at edges when employing the proposed method over VSRS. The obtained results using the computer generated image demonstrate negligible visual difference between the proposed method and VSRS. Fig. 4 (c) and (d) depict the virtual view and its difference to the ground truth.

6. ANALYSIS

Major artifacts due to DIBR mentioned in earlier works [4], are summarized in Table I, along with the applied remedies in VSRS and the proposed method.

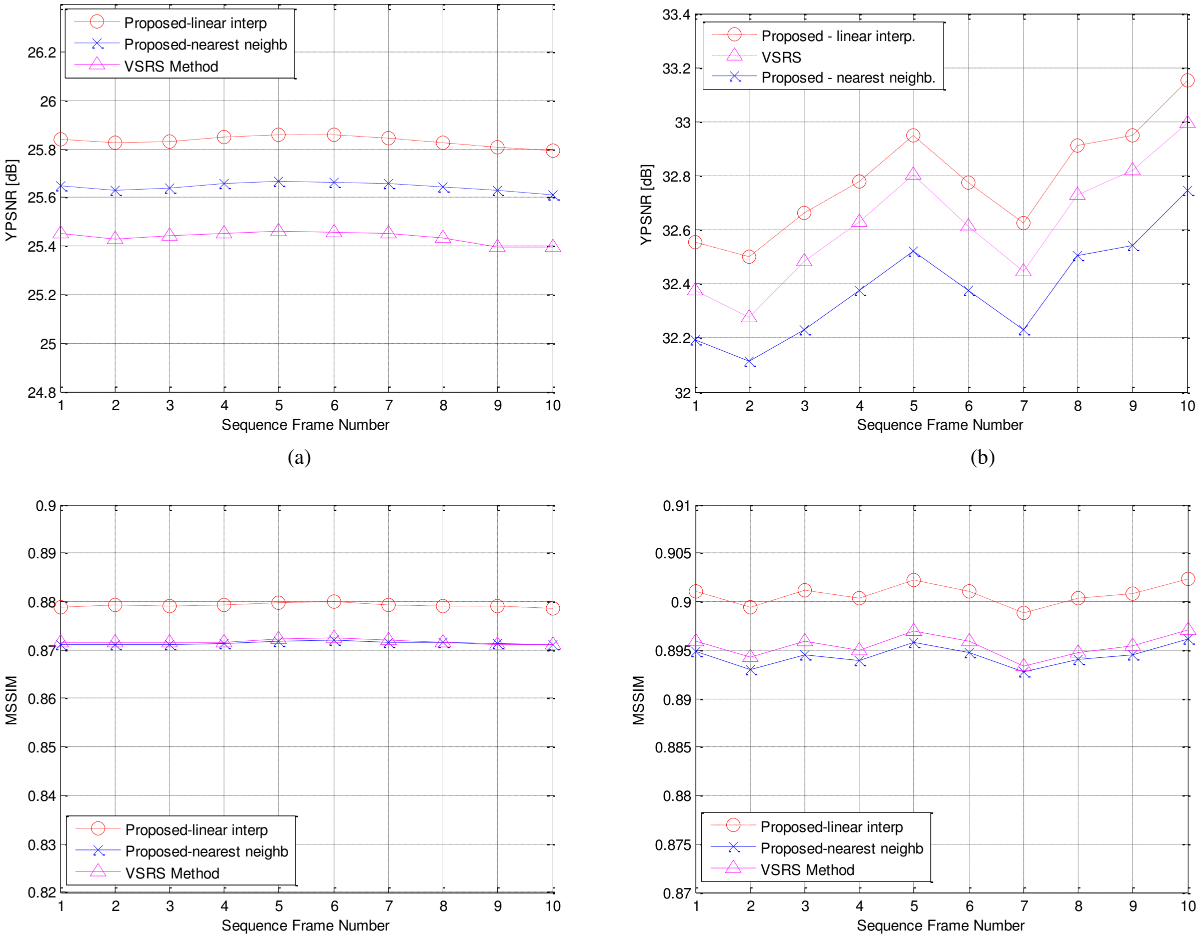
Empty cracks are a consequence of assigning each warped pixel to integer coordinates in the virtual view. In most DIBR methods, these cracks are found and filled by median filter. For the proposed method, these cracks do not appear because the floating point coordinates of rendered pixels are temporally stored and the values at integer coordinates are then interpolated. Therefore, interpolation is a simple way to avoid empty cracks.

Translucent cracks occur for the same reasons as empty cracks but contain background color as there happens to be such information at that position. VSRS puts constraints on pixel mappings in order to remove these. The proposed method removes all hidden pixels directly in the rendering process. Thus, no translucent cracks will appear in the resulting virtual view.

Empty regions are simply disoccluded areas as a consequence of the warping. VSRS propagates the background to fill these regions. In the proposed method, this is resolved by the introduced edge-pixels and interpolation. In fact, the effect is equivalent to background propagation for empty regions.

Smearing of edges appears due to smooth texture edges or depth-texture misalignments. (In DIBR using MVD sequences, this corresponds to Corona-like effects.) VSRS simply removes original pixels that would create this kind of artifact and fills the possible empty region in the virtual view by background propagation. The proposed method resolved this issue by selecting pure background and foreground color values for the edge-pixels in the original view. The proposed method selects the color at 1-2 pixels from each side of the edge in the original view based on the fact that each pixel integrates all light rays over the pixel surface at capture. Hence, a pixel may be a combination of both background and foreground colors at edges, incorporating more or less of each component depending on how much of each reaches that particular pixel. Therefore, the pure background and foreground color should in principle be at most one pixel away. Shading and other edge effects may however increase that distance.

Unnatural contours are the pixilation of "new" edges between background and foreground in the virtual view. They are the consequence of a sharp edge between foreground and background where colors of pixels at the edge have not been blended. VSRS employs splatting along edges in order to remove this jagged appearance. This process is the re-creation of averaged color values for edge pixels (described under Smearing of edges above),



(a)

(b)

(c)

(d)

Fig. 3. Objective metrics PSNR and SSIM the photographic sequences; (a) PSNR for each rendered frame at view position 4 of Lovebird; (b) PSNR for each rendered frame at view position 5 of Poznan Street; (c) MSSIM for each rendered frame at view position 4of Lovebird; (d) MSSIM for each rendered frame at view position 5 of Poznan Street.

which gives a much more natural appearance. We chose to smooth the edges by a combination of bicubic interpolation and an averaging filter.

In summary, the introduction of edge-pixels and an interpolation to assign values to the pixel grid of the virtual view have led to a straightforward algorithm that circumvent several special steps in other DIBR methods. The proposed method employs a one-dimensional interpolation, as a pure horizontal warping is applicable for distribution of 3DTV. The extension to a more general warping with arbitrary rotation would require a two-dimensional interpolation process. Such a procedure implies more computational complexity because the projected pixels have arbitrary locations and they need, therefore, to be sought for in a two-dimensional neighborhood of the virtual view pixels.

7. CONCLUSION

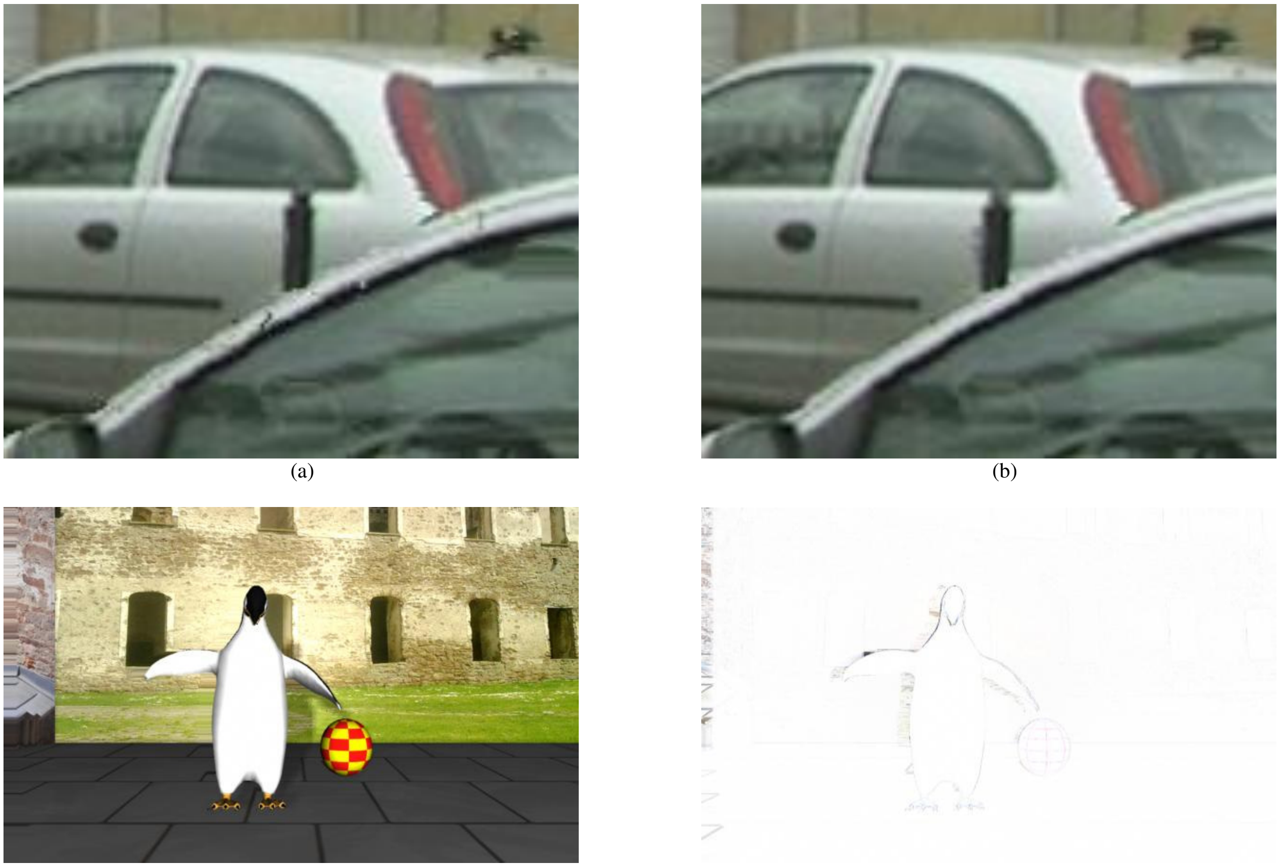
The present work investigated depth-image-based-rendering (DIBR). In particular it proposed a method that relies on fundamental principles: forward warping and interpolation in the rendered view. The artifacts as consequences of the 3D warping were studied. Their remedies in the DIBR implementation VSRS and the proposed method were analyzed. It was shown that the proposed method reduced most inherent artifacts as consequence of the introduced edge-pixels and the interpolation so that specific solutions are not required for each separate artifact. Furthermore, the proposed method demonstrated better quality when edges in the rendered view were smoothed by low-pass filtering. Future work will address temporal effects of the proposed method.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] 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, SPIE, pp. 93-104, 2004.



(a)

(b)

(c)

(d)

Fig. 4. Rendered virtual views using photographic “Poznan Street” sequence and computer generated image “Basketball Penguin” (a) VSRS; (b and c) Proposed method with linear interpolation; (d) Difference image to ground truth.

[2] O. Schreer et al., “3D Video communication: Algorithms, concepts and real-time systems in human centered communication,” John Wiley & Sons Ltd., 2005.

[3] C. L. Zitnick, et al., “High-quality video view interpolation using a layered representation,” ACM Trans. on Graphics, vol. 23, pp. 600-608, 2004.

[4] K. Müller, et al., “View Synthesis for Advanced 3D Video Systems,” EURASIP Journal on Image and Video Processing, Article ID 438148, 2008.

[5] L. Zhang et al., “Stereoscopic image generation based on depth images for 3D TV”, IEEE Transactions on Broadcasting, vol. 51, no. 2, pp. 191-199, Jun. 2005.

[6] “Report on experimental framework for 3D video coding,” ISO/IEC JTC1/SC29/WG11 MPEG2010/N11631, Guangzhou, China, October 2010.

[7] Domañski, et al., “Poznañ multiview video test sequences and camera parameters.” ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, Xian, China, 2009.

[8] Z. Wang, et al., “Image quality assessment: From error visibility to structural similarity,” IEEE Transactions on Image Processing vol 13, pp. 600 – 612, 2004.