A wavelet-based image quality metric for the assessment of 3D synthesized views

— Source link ☒

Emilie Bosc, Federica Battisti, Marco Carli, Patrick Le Callet

Related papers:

- Image quality assessment: from error visibility to structural similarity

- Evaluating virtual image quality using the side-views information fusion and depth maps

- Full Reference Quality Assessment of DIBR-Based Synthesized Images

- Perceptual quality assessment of 3D synthesized images

- NIQSV+: A No-Reference Synthesized View Quality Assessment Metric

A wavelet-based image quality metric for the assessment of 3D synthesized views

Emilie Bosc, Federica Battisti, Marco Carli, Patrick Le Callet

Applied Electronics Department
Università degli Studi Roma TRE
Rome, Italy

Polytech' Nantes

1. ABSTRACT

In this paper we present a novel image quality assessment technique for evaluating virtual synthesized views in the context of multi-view video. In particular, Free Viewpoint Videos are generated from uncompressed color views and their compressed associated depth maps by means of the View Synthesis Reference Software, provided by MPEG. Prior to the synthesis step, the original depth maps are encoded with different coding algorithms thus leading to the creation of additional artifacts in the synthesized views. The core of proposed wavelet-based metric is in the registration procedure performed to align the synthesized view and the original one, and in the skin detection that has been applied considering that the same distortion is more annoying if visible on human subjects rather than on other parts of the scene. The effectiveness of the metric is evaluated by analyzing the correlation of the scores obtained with the proposed metric with Mean Opinion Scores collected by means of subjective tests. These results are also compared against those of well known objective quality metrics. The experimental results confirm the effectiveness of the proposed metric.

Keywords: Image Quality Metric, 3D Video Quality, Depth Image Based Rendering

2. INTRODUCTION

Three dimensional (3D) imaging still greatly arises the public interest and curiosity as the wide range of proposed applications in the 3D market shows. Among others, application areas have been identified in on-line visualization for mixed reality environments[1–3] , communications[4, 5] , production or pre-visualization in television[6, 7] , games[8] , 3DTV[9, 10] , and education[11–13] .

3D systems should be able to provide an added-value compared to conventional monoscopic imaging (i.e. depth feeling or parallax motion) coupled with high image quality contents in order to ensure the success of the 3D imaging market. Quality assessment tools are thus needed for ensuring launches of viable products based on reliable quality evaluations. Image quality, depth quantity and visual comfort are three concepts involved in the estimation of the overall 3D Quality of Experience as assumed in[14] .

3D video currently involves considerable task forces in order to develop reliable assessment tools. Indeed, the success of 3D video applications[15] , namely 3D Television (3D TV) or Free Viewpoint Video (FVV), depends on the ability of 3D systems to provide high quality content. In particular, 3D video applications require the generation of novel virtual intermediate viewpoints to offer stereo- compliant pairs ensuring depth feeling, orto enable free navigation within the video scene. This is possible thanks to Depth-Image-Based-Rendering (DIBR)[16] algorithms that provide new viewpoints from acquired color views and associated depth maps of the scene as described in Section 3.

In this paper, we focus on the estimation of the image quality of synthesized frames, as a first step for assessing the quality of 3D contents. We believe that the evaluation of synthesized views quality is of tremendous importance, due to its use in most of the 3D imaging applications. Despite the advancements in view modeling

and synthesis, much less effort has been focused on developing algorithms for assessing the visual quality of a synthesized view. It is well known that the use of subjective tests is expensive, time consuming and the collected results may be affected by many factors which can significantly infer its reliability as loss of concentration of users during the test. The definition of an objective quality assessment metric for synthesized views is thus of paramount importance. As it has been observed in[17] , most of existing objective metrics are not well adapted to assess the quality of the virtual views. Artifacts related to DIBR systems are mostly located in specific areas that are the disoccluded regions. They are not scattered in the entire image such as specific 2-D video compression distortions. Consequently, as shown in[17] , commonly used 2-D quality metrics, that were originally designed to address 2-D video compression distortions, are not sufficient enough to assess the visual quality of synthesized views. They are able to indicate the presence of errors but not the degree of visual annoyance. A novel assessment tool can either improve the existing 2-D metrics, or propose a complete new approach.

Yasakethu et al.[18] proposed an adapted VQM (Video Quality Metric[19]) for measuring 3-D Video quality. It combines 2-D color information quality and depth information quality. Depth quality measurement is based on the analysis of the depth planes distortion. Results show higher correlation to subjective scores than the VQM. Another approach improving the existing 2-D metrics is the method proposed by Ekmekcioglu et al.[20] This depth-based perceptual tool can be applied to PSNR and SSIM. The method uses a weighting function based on depth data at the target viewpoint, and a temporal consistency function to take the motion activity into account. The study[20] shows that the proposed method enhances the correlation of PSNR and SSIM to subjective scores.

The method proposed in this paper is based on two main assumptions. Firstly, we believe that shifts can be introduced by the rendering process without affecting the visual quality of the synthesized images. And yet this type of artifact is penalized by pixel-by-pixel based quality metrics such as PSNR. Secondly, we believe that human beings are more sensitive to artifacts occurring around human being representations, which leads to severe quality scores. Our proposed metric is thus based on the consideration of shifting in the synthesized view and includes a skin detection based weighting.

We propose a new assessment tool for the objective quality evaluation of DIBR-based synthesized views. The proposed metric performs the Discrete Wavelet Transform (DWT) of the original and the synthesized views and, in this domain, the distortion present in the synthesized view is evaluated after a registration procedure used to align the content of the two images and by assigning a higher distortion score to distorted areas containing skin portions, relying on the assumption that an observer is more sensitive to distortions visible on human subjects rather than on other parts of the scene. In this study, we investigate the performances of our proposed metric, as well as that of state-of-the-art 2D metrics, by analyzing views synthesized from decompressed data. Our database contains FVV sequences generated through uncompressed color views and their associated compressed depth maps, from five different Multi-view Video plus Depth (MVD) sequences. Three codecs were used for depth maps coding, among which HEVC 6.1, H.264/AVC and JPEG 2000. Our FVV sequences database has been subjectively evaluated by 27 naïve observers. The effectiveness of the metric is assessed by computing the correlation between the scores obtained with the proposed metric and the Mean Opinion Scores collected by means of subjective tests. These results are also compared against that of well known objective quality metrics (i.e. PSNR, SSIM, MSSIM). The obtained results show that the proposed metric shows high correlation with human perception and better performances compared to state-of-the-art 2D quality metrics.

The rest of the paper is organized as follows: Section 3 shortly describes the View Synthesis Process and the problems this procedure can cause. Section 4 details the proposed algorithm and its effectiveness is analyzed in Section 5. Finally in Section 6 the conclusions are drawn.

3. VIEW SYNTHESIS PROCESS

As mentioned in the Introduction, 3D content can be represented by means of MVD video data: this technique is based on the use of a set of conventional color videos acquired at slightly different viewpoints and their associated depth maps. Newly generated virtual viewpoints by means of DIBR from MVD data are expected to confer as much comfort and quality as original acquired views.

A synthesized view is usually obtained by applying a homographic transformation (invertible transformation from a projective space to itself that maps straight lines to straight lines) to other views[21].

The homographic kernels can be obtained by estimating pixels correspondences between views at different depth values[22].

More in details, let us define with $I_{c_1}^t(x,y)$, $I_{c_2}^t(x,y)$, the intensity of a pixel acquired at time t by the camera c_1 and c_2 respectively, and with $I_{c_s}^t(x,y)$ the intensity of the corresponding synthesized view. Let us also define with $H_{c_1,c_v}[Z]$ and $H_{c_2,c_v}[Z]$ the homographic matrices at depth Z from view c_1 and c_2 to the synthesized view. The virtual view synthesis, at time t, can be expressed as:

$$I_{c_s}^t(x,y) = \begin{cases} w_{c_1}^t I_{c_1}^t(x,y) + w_{c_2}^t I_{c_2}^t(x,y) & \text{if } I_{c_s}^t(x,y) \text{ appears in both } c_1{}^t \text{ and } c_2{}^t \\ I_{c_1}^t(x,y) & \text{if } I_{c_s}^t(x,y) \text{ appears only in} c_1{}^t \\ I_{c_2}^t(x,y) & \text{if } I_{c_s}^t(x,y) \text{ appears only in } c_2{}^t \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

where $I_{c_1}^t(x,y)$, $I_{c_2}^t(x,y)$, and $I_{c_s}^t(x,y)$ are the pixel values of the matching points in different views, and $w_{c_1}^t$ and $w_{c_2}^t$ are the distance-dependent weights with $w_{c_1}^t + w_{c_2}^t = 1$.

This framework can be easily generalized to an acquisition system based on several cameras. In a real scenario, the number of available cameras is limited: to this aim many techniques exploiting views interpolation or warping have been proposed, as largely detailed in Section 2. However, these systems are prone to errors, mainly due to two factors: occlusions and depth estimation errors. In more details, *holes* in the synthesized view occur when the virtual viewpoint uncovers points that are invisible in the existing views. These artifacts can be due to the sampling process in the forward mapping process or to the real holes caused by occlusion boundaries in the disparity map. The first type of hole occurs when the disparity difference between adjacent pixels, is amplified in the mapping process. In the second case, the disparity difference is due to depth discontinuities.

DIBR techniques can be useful for filling the created disocclusions but they induce typical distortions. These distortions are different from those commonly encountered in 2D video compression, such as blur or blocking artifacts. In fact, due to the warping process, DIBR algorithms induce geometric distortions mainly located around object boundaries. Moreover, depending on the in-painting method used by the DIBR algorithm, the types of artifacts may also differ. For these reasons, commonly used 2D image quality metrics fail in rendering quality scores correlated to human perception and there is an increasing need for defining new quality metrics. Post processing of acquired video sequences is often necessary to provide a good 3D Quality of Experience[23]. As explained in[23], scene depth adaptation and 3D display adaptation among others may require the use of view synthesis. In addition, since the 3D processing chain expects MVD data to be compressed before generating new virtual viewpoints from the decompressed data, it is of paramount importance to address the issue of the objective assessment of views synthesized from decompressed data. In other words, the combination of compression related artifacts and synthesis related artifacts needs to be evaluated.

The problem of generating views arises in the multi-view data coding chain. In this case, for reducing the requested resources, alternative representations of original data have been proposed. Video-plus-depth representation[24] is based on a regular video stream, where each frame is enriched with a depth map providing a Z-value for each pixel. The final left and right views are reconstructed by DIBR. The video-plus-depth representation can be evolutionary built on the existing DVB infrastructure and is considered as the most broadcast-friendly representation. ISO/IEC JTC1/SC29/WG11 (MPEG) issued a Call for Proposals on 3D Video Coding in March 2011 for designing a new coding framework for MVD data and several proposals were suggested[25]. Video plus depth approaches allow view rendering with variable baseline and increase the data compression rate. However, several artifacts may be introduced through the rendering process. This scenario allows to better highlight the need for the definition of ad-hoc quality metrics.

4. METRIC DESCRIPTION

The proposed method is based on the analysis in the frequency domain of DIBR images based on the evidence that the introduced artifacts (holes) are mainly characterized by high frequencies in the horizontal direction. To

this aim image degradation is evaluated by measuring the Kolmogrov-Smirnov distance between the histograms of the horizontal details subband of the original view and of the synthesized one. As well known spatial localization is lost when dealing with histograms. Thus, a block based procedure is adopted in order to perform a spatial localization of the artifacts. Moreover DIBR images are often affected by a displacement between different views. For this reason a registration procedure is employed before proceeding to the distortion computation.

In the development of our metric we also considered the fact that the same artifact affecting image regions with and without human beings are perceived differently. In particular, in the first case the artifacts are more annoying. To cope with this task a skin detection is performed in order to *perceptually* weight each image block containing artifacts on human subjects. In more details, the adopted skin detector is based on a color segmentation in the HSV color space[26] . In particular, the H channel is used to select a color range ([0.064 - 0.085] in our case) for performing skin detection.

In the following, the details of the proposed objective image quality metric for DIBR images are reported and are illustrated in Figure 1.
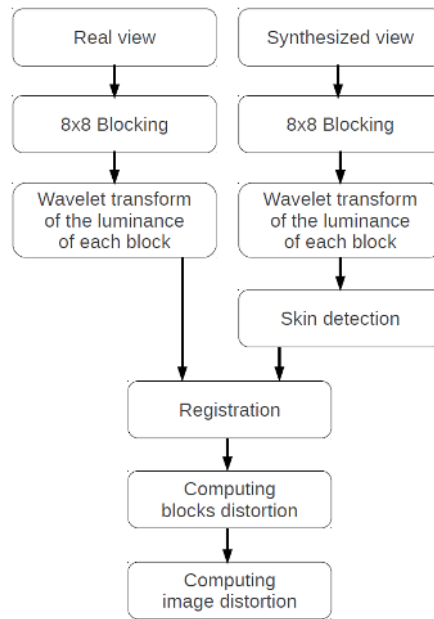

Figure 1. Block diagram of the proposed metric.

As previously mentioned, the first step in the proposed metric consists in eliminating the displacement between the two views. In order to do so, a simple block matching algorithm is run in order to register corresponding blocks before quality evaluation.

The following step consists in analyzing histograms variations in the horizontal details subband. To this aim the Kolmogorov-Smirnov test (K-S test) is adopted for quantifying the distance between the empirical distribution function of a sample and the cumulative distribution function of the reference distribution. The test statistic is computed as:

$$\max\left(\left|F_1\left(x\right) - F_2\left(x\right)\right|\right) \tag{2}$$

where $F_1\left(x\right)$ and $F_2\left(x\right)$ are respectively the cumulative distribution functions.

This distance value is used for computing the block distortion d_b:

$$d_b = \max\left(\left|F_p\left(x\right) - F_q\left(x\right)\right|\right) * w_s \tag{3}$$

where $F_p(x)$ and $F_q(x)$ are the cumulative distribution functions of the histograms p and q of the horizontal details subband computed on the luminance component of the generic block b of the original image and of the synthesized one respectively and w_s is the weighting factor that takes into account the presence of the skin.

The block distortion d_b with $b = 1, ..., B$ is used for evaluating the overall image distortion as:

$$d = \frac{1}{D_0} \sum_{b=1}^{B} d_b \tag{4}$$

where D_0 is a normalization constant.

Finally, the image quality score for each frame is computed as:

$$s = \frac{1}{1+d} = \frac{1}{1 + \frac{1}{D_0} \sum_{b=1}^{B} \max(|F_p(x) - F_q(x)|)} \tag{5}$$

In order to obtain the overall video score, the Minkowski formula is applied:

$$s = \left(\frac{1}{n} \sum_{i=1}^{n} s_i^{\beta} \right)^{1/\beta} \tag{6}$$

where s_i is the ith frame score, n is the number of frames and $\beta = 322$.

The score range is 0 (when $d \to \infty$) to 1 (when $d = 0$). In fact, a lower distortion corresponds to a larger score and a higher distortion corresponds to a lower score.

5. VALIDATION OF THE PROPOSED METRIC

The goal of this work is to propose an objective quality metric assessing the quality of DIBR-based synthesized views. Commonly used objective quality metrics already target the assessment of 2D compression related artifacts in conventional 2D imaging. As explained in Section 3, the special feature of the assessment of DIBR-based synthesized views lies in the fact that depth data is used to generate a novel view from projective geometry principles. Resulting artifacts are mainly located around object's edges, due to erroneous depth data or numerical rounding of pixel positions. Our proposed objective method targets this case of use. In order to evaluate the performances of the proposed quality metric we consider an experimental protocol involving the conditions of the targeted case of use. A set of objective quality metrics (including our proposed metric) are used to rate the quality of a database of synthesized views. This section first addresses the experimental protocol including the description of the used database of synthesized views. Then, in a second section, the results are discussed.

5.1 Experimental protocol

In these experiments we considered five sequences of 100 frames (and 10fps) of the database, built from 50 intermediate virtual synthesized frames that simulate a smooth camera motion from left to right and from right to left, similarly to a FVV context of use. Each synthesized frame of the sequences is generated from the corresponding original color view and decompressed depth data. Depth coders under test include HEVC 6.1 and H264 (JM 18) and JPEG2000, all in intra coding mode. For H.264, we used the JM 18.4 (Joint Multiview Video Model) software for the Multiview Video Coding (MVC) project of the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG)[?] . We believe that this experimental protocol provides different types of depth data compression-related artifacts resulting in different types of distortion in the synthesized views. Three test quantization parameters were selected for each depth codec under test according to the visual quality of the rendered views. This procedure was motivated by the need to cover a wide range of categories in the visual quality scale in order to properly define each codec under test. Table 2 gives the details of the quantization parameters used in these experiments. Six MVD sequences are used in these experiments: *Book Arrival*, *Newspaper*, *Kendo* and *Balloons* are real scenes; and *GT_Fly* and *Undo_Dancer*

Sequence Name	Resolution	No. of frames	Camera Arrangement
Book Arrival	1024×768	100	16 cameras with 6.5cm spacing
Balloons	1024×768	300	7 cameras with 5 cm spacing, moving camera array
Kendo	1024×768	300	7 cameras with 5 cm spacing, moving camera array
GT_Fly	1920×1080	250	Computer-generated imagery with ground truth depth data
Undo_Dancer	1920×1080	250	Computer-generated imagery with ground truth depth data

Table 1. Six MVD sequences used in the database.

Depth codec	Quantization parameters
H.264 (JM18)	Qp = [{Book Arrival, Balloons, Kendo, Newspaper}{25, 33, 47}, Undo Dancer{25,40,47}, Gt Fly{30,40,47}]
HEVC 6.1	Qp = [{All of the sequences}{34, 45, 50}]
JPEG2000	0.05bpp, 0.009bpp and 0.005bpp

Table 2. Quantization parameters applied on the depth data to generate the synthesized views of the database.

are synthetic scenes. Table 1 summarizes the features of the sequences. The sequences and the key frames were selected for their availability and amount of depth. Table 3 gives the details of the encoded viewpoints and the target viewpoint for the synthesis. The synthesis process is performed through the 3D-HTM 0.4 renderer, that is the view synthesis algorithm used in MPEG 3DV group of standardization at the time of writing this paper. We set the *Blended Mode* parameter of the synthesis algorithm for using the right view only for hole filling instead of carrying out a weighted average of samples extrapolated from both sides (as done in the MPEG evaluations).
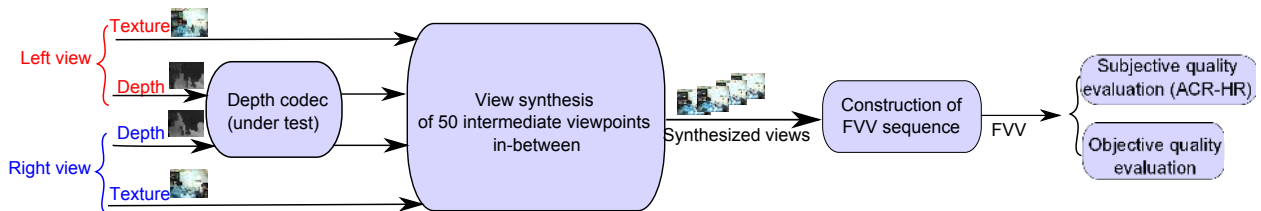

Figure 2. Overview of the experimental protocol.

5.1.1 Subjective quality assessment

Twenty-seven naive observers participated in the subjective quality evaluation test into two 30-minute sessions. ACR-HR (Absolute Category Rating with Hidden Reference)[27] methodology was used to assess 288 FVV sequences, among which were the forty-five hereby considered. ACR-HR methodology[27] consists in presenting each stimulus only once to the observers, who are asked to rate the quality of the stimuli relying on a five-level quality scale (5: *Excellent*; 4: *Good*; 3: *Fair*; 2: *Poor*; 1: *Bad*). The reference version of each stimulus is included in the test procedure and rated like any other stimulus. This is referred to as a "hidden reference

Sequence Name	Encoded viewpoints	Frame no.
Book Arrival	$10 - 6$	33
Balloons	$1 - 5$	1
Kendo	$1 - 5$	1
GT_Fly	$1 - 9$	157
Undo_Dancer	$1 - 9$	250

Table 3. Input and output views used to generate the synthesized views of the database.

condition". The subjective evaluations were conducted in an ITU conforming test environment. The stimuli were displayed on a Panasonic BT-3DL2550 screen (1920×1080p), and according to ITU-T BT.500[28] .

5.1.2 Objective quality assessment

Regarding the objective quality assessment, twelve objective quality metrics, including our proposal, rated the tested sequences. The objective measurements were realized by the means of MetriX MuX Visual Quality Assessment Package software[29] .

The reliability of the objective metrics can be evaluated through the calculation of the correlation between ACR-HR DMOS (Differential Mean Opinion Score) scores and objective quality metrics scores. In this paper, this is assessed through the computation of the Pearson Linear Correlation Coefficients (PLCC) for the whole fitted measured points. Fitted measure points correspond to so called predicted DMOS, noted as $DMOS_p$. Video Quality Expert Group (VQEG) Phase I FR-TV[30] recommends the use of the following logistic function for the fitting step:

$$DMOS_p = a.score^3 + b.score^2 + c.score + d \qquad (7)$$

where *score* is the obtained score from the objective metric and $\{a, b, c, d\}$ are the parameters of the cubic function. They are obtained through the regression step to minimize the difference between $DMOS$ and $DMOS_p$. The Pearson linear correlation coefficients are then computed though:

$$PLCC = \frac{\sum_{i=1}^{N} \left(DMOS_i - \overline{DMOS}\right) \left(DMOS_{p_i} - \overline{DMOS_p}\right)}{\sqrt{\sum_{i=1}^{N} \left(DMOS_i - \overline{DMOS}\right)^2} \sqrt{\left(DMOS_{p_i} - \overline{DMOS_p}\right)^2}} \qquad (8)$$

where $\overline{DMOS}$ and $\overline{DMOS_p}$ are the average of $DMOS$ and $DMOS_p$ over the N stimuli.

Table 4 gives the results. This table gives the PLCC, the outlier ratio and the Root Mean Square (RMSE). Our proposed metric obtains the highest correlation score with the subjective scores (90.1%) and also the minimum outlier ratio (0) and RMSE scores (0.46). Out of the 2D usual metrics, IFC is the one obtaining the highest correlation score (87%), in these experiments. These results show that our proposed metric outperforms the 2D usual objective quality metric when assessing the quality of views synthesized from compressed data. Indeed, these experiments included three different coding methods whose impact in the synthesized views is different.

6. CONCLUSIONS

In this paper, a novel image quality assessment method for the quality evaluation of views synthesized from decompressed data is presented. The proposed method is based on wavelet decomposition of the views and exploits a registration and a skin detection step. The validation experiments included synthesized views generated from decompressed depth data. Three different coding methods where used to encode the depth data and thus induced different types of artifacts. Then the view synthesis process also induces typical distortions. Our proposed metric showed 90% correlation with the subjective scores and outperformed the tested 2D usual objective quality metrics. In the future, we plan to test the performances of our proposed metrics in other contexts of use and to extend it to 3D stereoscopic quality evaluation.

	PLCC (%)	Out-lier ratio	RMSE
Proposed method	**90.1**	**0**	**0.46**
PSNR	73.3	**0**	0.73
SNR	63	**0**	0.84
MSE	55.5	**0**	0.89
SSIM	58.2	**0**	0.87
MSSIM	43.7	0.02	0.57
VSNR	35	3.6	0.65
VIF	59.5	0.04	0.86
VIFP	54.9	0.02	0.89
UQI	73.4	**0**	0.73
IFC	87	**0**	0.52
NQM	36.3	0.02	1
WSNR	55	0.02	0.89

Table 4. Pearson correlation coefficients (PLCC) between DMOS and objective scores in percentage, outlier ratio values and RMSE.

REFERENCES

[1] Allard, J., Franco, J., Menier, C., Boyer, E., and Raffin, B., "The grimage platform: A mixed reality environment for interactions," in [*Proc. IEEE Int. Conf. Computer Vision Systems ICVS '06*], (2006).

[2] Kulik, A., Kunert, A., Beck, S., Reichel, R., Blach, R., Zink, A., and Froehlich, B., "C1x6: a stereoscopic six-user display for co-located collaboration in shared virtual environments," in [*Proceedings of the 2011 SIGGRAPH Asia Conference*], *SA '11*, 188:1–188:12, ACM, New York, NY, USA (2011).

[3] Theoktisto, V. and Fairén, M., "Enhancing collaboration in virtual reality applications," *Comput. Graph.* **29**, 704–718 (Oct. 2005).

[4] Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., Koller-Meier, E., Svoboda, T., Van Gool, L., Lang, S., Strehlke, K., Moere, A. V., and Staadt, O., "blue-c: a spatially immersive display and 3d video portal for telepresence," in [*ACM SIGGRAPH 2003 Papers*], *SIGGRAPH '03*, 819–827, ACM, New York, NY, USA (2003).

[5] Schroeder, R., "Being there together and the future of connected presence," *Presence: Teleoper. Virtual Environ.* **15**, 438–454 (Aug. 2006).

[6] Zilly, F., Muller, M., Eisert, P., and Kauff, P., "The stereoscopic analyzer - an image-based assistance tool for stereo shooting and 3D production," in [*Image Processing (ICIP), 2010 17th IEEE International Conference on*], 4029 –4032 (Sept. 2010).

[7] Zilly, F., Muller, M., Kauff, P., and Schafer, R., "STAN - an assistance system for 3d productions: From bad stereo to good stereo," in [*Electronic Media Technology (CEMT), 2011 14th ITG Conference on*], 1 –6 (march 2011).

[8] Starck, J. and Hilton, A., "Surface capture for performance-based animation," *Computer Graphics and Applications, IEEE* **27**, 21 –31 (may-june 2007).

[9] Nahrstedt, K., Yang, Z., Wu, W., Arefin, A., and Rivas, R., "Next generation session management for 3D teleimmersive interactive environments," *Multimedia Tools Appl.* **51**, 593–623 (Jan. 2011).

[10] Grau, O., Muller, M., and Kluger, J., "Tools for 3D-TV programme production," in [*3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*], 1 –4 (may 2011).

[11] Navarro, A., Pradilla, J., and Madrian, P., "A 3D game tool for mobile networks planning," in [*Mobile, Hybrid, and On-Line Learning, 2010. ELML '10. Second International Conference on*], 158 –161 (feb. 2010).

[12] Floryan, M. and Woolf, B., "Students that benefit from educational 3D games," in [*Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on*], 473 –475 (july 2011).

[13] Bedi, K., Hrustek, N., and Coric, A., "Teaching vs. 3D gaming in secondary school," in [*MIPRO, 2011 Proceedings of the 34th International Convention*], 1325 –1330 (may 2011).

[14] Chen, W., Fournier, J., Barkowsky, M., and Le Callet, P., "Quality of experience model for 3DTV," (Jan. 2012).

[15] Smolic, A., Mueller, K., Merkle, P., Fehn, C., Kauff, P., Eisert, P., and Wiegand, T., "3D video and free viewpoint VideoTechnologies, applications and MPEG standards," in [*Proceedings of the IEEE International Conference on Multimedia and Expo (ICME06)*], 2161–2164 (2006).

[16] Fehn, C., "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in [*Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI*], **5291**, 93–104 (2004).

[17] Bosc, E., Pepion, R., Le Callet, P., Koppel, M., Ndjiki-Nya, P., Pressigout, M., and Morin, L., "Towards a new quality metric for 3-d synthesized view assessment," *Selected Topics in Signal Processing, IEEE Journal of* **5**, 1332 –1343 (nov. 2011).

[18] Yasakethu, S. L. P., Worrall, S. T., De Silva, D., Fernando, W. A. C., and Kondoz, A. M., "A compound depth and image quality metric for measuring the effects of packet loss on 3D video," in [*Proc. of 17th International Conference on Digital Signal Processing*], (July 2011).

[19] "http://www.its.bldrdoc.gov/resources/video-quality-research/software.aspx."

[20] Ekmekcioglu, E., Worrall, S. T., De Silva, D., Fernando, W. A. C., and Kondoz, A. M., "Depth based perceptual quality assessment for synthesized camera viewpoints," in [*Proc. of Second International Conference on User Centric Media, UCMedia 2010*], (Sept. 2010).

[21] Hartley, R. and Zisserman, A., [*Multiple View Geometry in Computer Vision*], Cambridge University Press, 2 ed. (Apr. 2004).

[22] Liu, Y., Huang, Q., Ma, S., Zhao, D., and Gao, W., "Joint video/depth rate allocation for 3d video coding based on view synthesis distortion model," *Image Communication* **24**, 666–681 (Sept. 2009).

[23] Smolic, A., Kauff, P., Knorr, S., Hornung, A., Kunter, M., Muller, M., and Lang, M., "Three-dimensional video postproduction and processing," *Proceedings of the IEEE* **99**(4), 607–625 (2011).

[24] Fehn, C., Kauff, P., de Beeck, M. O., Ernst, F., IJsselsteijn, W., Pollefeys, M., Gool, L. V., Ofek, E., and Sexton, I., "An evolutionary and optimised approach on 3D-TV," in [*Proc. International Broadcast Conference*], 357–365 (2002).

[25] Schwarz, H., Bartnik, C., Bosse, S., Brust, H., Hinz, T., Lakshman, H., Marpe, D., Merkle, P., Muller, K., Rhee, H., Tech, G., Winken, M., and Wiegand, T., "3D video coding using advanced prediction, depth modeling, and encoder control methods," in [*Picture Coding Symposium (PCS), 2012*], 1 –4 (May 2012).

[26] Oliveira, V. and Conci, A., "Skin detection using HSV color space," in [*Sibgrapi 2009*], (2009).

[27] ITU-T, "Subjective video quality assessment methods for multimedia applications," Tech. Rep. Rec. P910, Geneva (2008).

[28] BT., I., [*500, Methodology for the subjective assessment of the quality of television pictures*], November (1993).

[29] "Metrix mux home page."

[30] Video Quality Experts Group, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *VQEG, Mar* (2000).