MEASUREMENT OF CRITICAL TEMPORAL INCONSISTENCY FOR QUALITY ASSESSMENT OF SYNTHESIZED VIDEO

*Hak Gu Kim and Yong Man Ro**

Image Video System Lab, School of Electrical Engineering, KAIST, Republic of Korea

ABSTRACT

This paper proposes a new temporal consistency measure for quality assessment of synthesized video. Disocclusion regions appear hole regions of the synthesized video at virtual viewpoints. Filling hole regions could be problematic when the synthesized video is perceived through multi-view displays. In particular, the temporal inconsistency caused by hole filling process in view synthesis could affect the perceptual quality of the synthesized video. In the proposed method, we extract excessive flicker regions between consecutive frames and quantify the perceptual effects of the temporal inconsistency on them by measuring the structural similarity. We have demonstrated the validity of the proposed quality measure by comparisons of subjective ratings and existing objective metrics. Experimental results have shown that the proposed temporal inconsistency measure is highly correlated with the overall quality of the synthesized video.

Index Terms— Temporal inconsistency measure, synthesized video, view synthesis, visual quality assessment

1. INTRODUCTION

In recent years, multi-view imaging systems such as autostereoscopic display and free-viewpoint TV have drawn plenty of attention of video industry and customers [1], [2]. In general, the multi-view imaging systems need a large number of views at different virtual viewpoints to provide multiple perspectives of the same scene. As a result, view synthesis is required to support the multi-view imaging systems by generating virtual views at desired viewpoints [3].

Depth image based rendering (DIBR) is a widely used view synthesis technique. The DIBR contains 3D warping and hole filling techniques [4], [5]. In 3D warping, a warped video is generated by mapping a given reference color video into desired virtual viewpoints with the associated depth video. In this process, hole regions could appear in the warped video since backgrounds occluded by foregrounds could be seen. Hole filling comes next to fill the hole regions of the warped video [6]-[9].

The existing methods based on DIBR, however, could not provide satisfactory quality for synthesized video [10], [11]. Most existing view synthesis methods did not consider the characteristics of human visual perception for providing better viewing quality of synthesized video [12]-[16]. As such, synthesized video could contain visible artifacts like temporal inconsistency. Visible artifacts in synthesized video even cause perceptual problems (e.g., disturbing flicker artifacts) in human visual systems. To cope with the quality issues of view synthesis, a reliable objective quality assessment for synthesized video is needed.

Most of earlier studies tried to devise the quality assessment metrics based on existing 2D quality assessment methods such as PSNR and SSIM for 3D image/video [17], [18]. They focused on the quality of stereoscopic 3D (S3D) image/video rather than that of multiple synthesized videos generated by view synthesis. The distortions of synthesized video such as geometric distortions and temporal inconsistency could not be observed in the conventional S3D image/video. These distortions might be caused by view synthesis algorithms [10], [19], [20]. Therefore, the existing quality assessment based objective metrics might be unreliable to deal with the quality of the synthesized video.

Unlike the 2D video quality assessment [29]-[31], a few full reference quality assessments for the synthesized views were proposed. Conze *et al*. devised an objective metric considering textures, gradient orientations, and contrasts of the synthesized views [19]. In [20], Battisti *et al*. proposed a 3DSwIM using the differences between histograms of the synthesized views and reference views in wavelet domain. However, for the synthesized videos in multi-view imaging, the full reference quality assessment approaches are not appropriate. In multi-view imaging systems, it is not possible to provide corresponding reference videos at multiple virtual viewpoints. In addition, most existing studies were limited in spatial distortions of the synthesized views without considering the visible artifacts in time domain [32].

In this paper, we investigate temporal inconsistency in order to assess the quality of synthesized video distorted by view synthesis. In synthesized videos, severe visible artifacts caused by view synthesis could mainly occur around specific regions (e.g., hole regions) [10]. So quality assessment

* Corresponding author (ymro@ee.kaist.ac.kr)

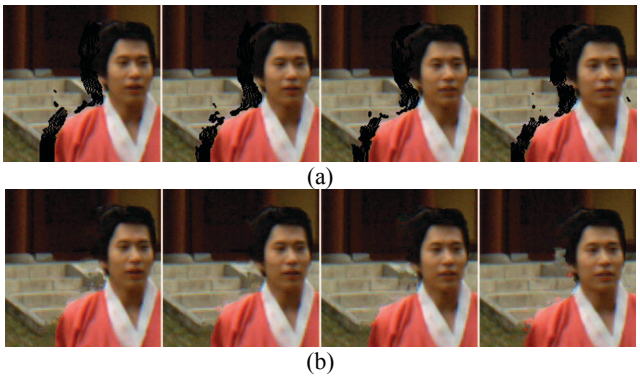

Fig. 1. Example of temporal inconsistency over consecutive frames for "Lovebird1" (143-146th frames). (a) Consecutive frames with hole regions. (b) Hole-filled results by view synthesis based on DIBR [7]. Black color regions in (a) represent hole regions.

performance became poor when existing quality metrics were applied on the entire region [20], [21]. From these observations, we propose a new critical temporal inconsistency (CTI) measure to effectively and objectively assess the quality of the synthesized video by measuring the structural similarity only on the specific regions. To detect the specific regions, we calculate the differences between temporally neighboring frames. Then, a threshold is applied to select the most important errors. In this literature, these regions are identified as excessive flicker regions. Finally, we measure the structural similarity only on excessive flicker regions in synthesized video. In the proposed method, by explicitly measuring the structural similarity between consecutive frames of the synthesized video, reference video are not required to quantify the quality of the synthesized video at virtual viewpoints. Hence, the proposed objective measurement is more applicable than existing full reference methods in multi-view imaging systems. Experimental results show that the proposed CTI index has high correlation with human subjective scores.

The remainder of this paper is organized as follows. In Section 2, we describe the proposed temporal inconsistency measurement method to objectively assess the quality of the synthesized video. In Section 3, we describe experiments and results. Finally, conclusions are drawn in Section 4.

2. PROPOSED METHOD

In this section, we present the proposed measurement for quantifying the perceptual effects of the disturbing flicker artifacts of the synthesized video. Fig. 1 shows an example of temporal inconsistency on synthesized video. As is seen in Fig. 1, structural distortions around hole regions are changing in consecutive frames. To properly measure overall quality of synthesized video, in this paper, we take into account the structural distortions not on entire region but around specific regions of synthesized video. In the next subsections, we describe the proposed temporal inconsistency measurement to objectively assess the quality of synthesized video.

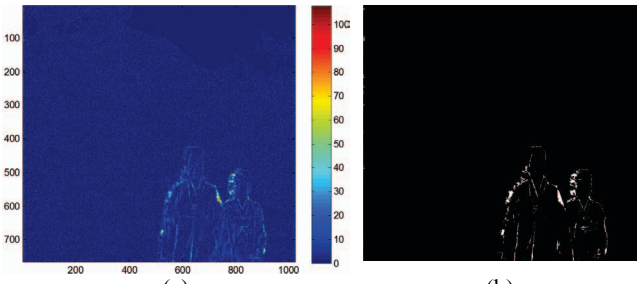

Fig. 2. Example of excessive flicker regions detection for "Lovebird1" (134-135 frames). (a) Motion-compensated difference map between the consecutive frames generated by DIBR method [7]. (b) Excessive flicker region map. In (b), white pixels represent the excessive flicker regions.

2.1. Excessive flicker regions detection

In this section, we describe the specific regions (denoted as excessive flicker regions) that are highly likely to induce temporal inconsistency (e.g., flicker) in the synthesized video. The excessive flicker regions indicate the differences between temporally neighboring frames of synthesized video, whose effect causes perceptual problem in human visual system.

In the proposed method, motion-compensated differences between temporally neighboring frames are found to avoid including differences due to object motions. It can be written as

$$\mathbf{E}_t(x,y) = |\,\mathbf{I}_t(x,y) - \hat{\mathbf{I}}_t(x,y)\,|, \tag{1}$$

where $\mathbf{E}_t$ represents the motion-compensated difference map. $\mathbf{I}_t$ denotes the t-th synthesized frame and $\hat{\mathbf{I}}_t$ denotes the motion compensated frame obtained by the backward warping from the (t-1)-th synthesized frame. It can be written as

$$\hat{\mathbf{I}}_t(x+\mathbf{MV}_t^h(x,y), y+\mathbf{MV}_t^v(x,y)) = \mathbf{I}_{t-1}(x,y), \tag{2}$$

where $\mathbf{I}_{t-1}$ denotes the (t-1)-th synthesized frame. $\mathbf{MV}_t^h$ and $\mathbf{MV}_t^v$ represent horizontal and vertical motion vectors at t-th frame, respectively. Motion vectors are obtained by an optical flow method [22].

By using Eq. (1), the motion-compensated differences are obtained between the corresponding regions in temporally neighboring frames. Fig. 2(a) shows an example of the motion-compensated differences.

Among these differences, the excessive flicker regions are extracted with a threshold (Th_t) which can be written as,

$$Th_t = \frac{\max\{\mathbf{E}_t(x,y) \mid x \in 1,\cdots,W.\ y \in 1,\cdots,H\}}{C_1}, \tag{3}$$

where $\max\{\}$ is a maximum operator. W and H represent the

width and height of image, respectively. The constant C_1 is to adjust the percentage of the maximum differences for detecting excessive flicker regions in each frame. C_1 is 10 which has been found experimentally. The used threshold in this paper indicates 10 percent of maximum difference.

Finally, an excessive flicker region mask ($\mathbf{M}_t$) is obtained by applying the threshold Th_t, with which excessive flicker regions are segmented. This mask can be written as

$$\mathbf{M}_t(x,y) = \begin{cases} 1, & if\ \mathbf{E}_t(x,y) \geq Th_t \\ 0, & otherwise \end{cases}, \quad (4)$$

where $\mathbf{M}_t$ denotes the excessive flicker region mask at the t-th frame, which is one for the differences above Th_t and zero otherwise.

Fig. 2(b) shows an excessive flicker region map. In Fig. 2(b), white pixels indicate excessive flicker regions. As shown in Fig. 2(b), the pixels on excessive flicker regions are mainly located around specific regions such as hole regions or object boundaries.

2.2. Structural similarity on excessive flicker regions

Structural similarity on the excessive flicker regions between the temporally neighboring frames is measured in order to quantify the perceptual effects of temporal inconsistency. To maintain temporal consistency in the synthesized video, structural mismatches on the corresponding regions are supposed to be minimized.

To measure structural similarities on pixels belonging to excessive flicker regions, a widely used quality metric SSIM (Structural Similarity [23]) is adopted. Let $\mathbf{\Omega}$ denote a set of pixels in excessive flicker regions. (i.e., $\mathbf{\Omega} = \{(x,y) \mid \mathbf{M}_t(x,y) = 1\}$. The structural similarity on excessive flicker regions between temporally neighboring frames is denoted as CTI index in this paper. CTI index for the t-th can be written as

$$CTI_t(\mathbf{I}_t, \hat{\mathbf{I}}_t, \mathbf{M}_t) = \frac{1}{|\mathbf{\Omega}|} \sum_{(x,y)\in\mathbf{\Omega}} SSIM(\boldsymbol{l}_{x,y}, \hat{\boldsymbol{l}}_{x,y}), \quad (5)$$

where $|\mathbf{\Omega}|$ denotes the number of pixels in $\mathbf{\Omega}$. $\boldsymbol{l}_{x,y}$ and $\hat{\boldsymbol{l}}_{x,y}$ represent local windows (size of 11 x 11) centered at (x,y) in $\mathbf{I}_t$ and $\hat{\mathbf{I}}_t$, respectively. Note that 11 x 11 window yielded a good result of local image statistics in [23].

In Eq. (5), SSIM index between the two local windows $\boldsymbol{l}_{x,y}$ and $\hat{\boldsymbol{l}}_{x,y}$, can be written as

$$SSIM(\boldsymbol{l}_{x,y}, \hat{\boldsymbol{l}}_{x,y}) = \frac{(2\mu_l\mu_{\hat{l}} + C_2)(2\sigma_{l\hat{l}} + C_3)}{(\mu_l^2 + \mu_{\hat{l}}^2 + C_2)(\sigma_l^2 + \sigma_{\hat{l}}^2 + C_3)}, \quad (6)$$

where μ_l and σ_l are mean value and standard deviation in the local window $\boldsymbol{l}$, respectively. $\sigma_{l\hat{l}}$ denotes the covariance between $\boldsymbol{l}$ and $\hat{\boldsymbol{l}}$. C_2 and C_3 are constants to avoid zero of the denominator. In our experiment, C_2=6.50 and C_3=58.52 as used in [23].

2.3. Temporal pooling

To obtain a final CTI score for synthesized video, temporal pooling for all CTI scores obtained from frames is needed. In this paper, a weighted average pooling is adopted. On the entire frames of the synthesized video, several frames could cause disturbing flicker artifacts by excessive structural mismatches between temporally neighboring frames. As a result, the degree of flicker artifacts is proportional to the number of pixels on the excessive flicker regions at each frame. Consequently, the final CTI score for the synthesized video can be written as

$$CTI = \sum_{t=1}^{T} \omega_t \times CTI_t, \quad (7)$$

where T represents the number of frames for synthesized video. ω_t is a weight value for the t-th synthesized frame. It can be written as

$$\omega_t = \frac{N_t}{\sum_{i=1}^{T} N_i}, \quad (8)$$

where N_t denotes the number of pixels belonging to the excessive flicker regions at the t-th synthesized frame. The weight decreases as the number of pixels on the excessive flicker regions decreases.

3. EXPERIMENTS AND RESULTS

To verify the proposed temporal inconsistency measure for objective quality assessment for synthesized video, experiments have been performed with DIBR video datasets. In next subsections, we describe datasets used in our experiments and performance of the proposed method.

3.1. Datasets

We used IRCCyN/IVC DIBR video datasets, which contained 84 synthesized videos (= 3 datasets x 4 virtual viewpoints x 7 view synthesis algorithms) of about 6 seconds in 1024 x 768 resolution [24]. The IRCCyN/IVC DIBR video datasets consist of three different multi-view plus depth sequences (MVD), which are Book Arrival, Lovebird1, and Newspaper. Table 1 illustrates a detail description of each dataset and view synthesis conditions.

In these database, four different synthesized videos were generated by DIBR for each dataset. In Table 1, '8→9' means that the reference video at *Cam* 8 was warped to the

Table 1. Datasets and view synthesis conditions in [24].

Dataset	Frame rate	Number of frames	Virtual viewpoints
Book Arrival	15Hz	100	8→9, 8→10, 10→8, 10→9
Lovebird1	30Hz	150	6→7, 6→8, 8→6, 8→7
Newspaper	30Hz	200	4→5, 4→6, 6→4, 6→5

virtual viewpoint at *Cam* 9 to generate synthesized video.

In the IRCCyN/IVC DIBR video datasets, test videos were generated by using seven different algorithms: A1) Fehn's method1 (borders are cropped) [3], A2) Fehn's method2 (borders are inpainted) [3], A3) Tanimoto's method [14], A4) Muller's method [15], A5) Ndjiki-Nya's method [7], A6) Koppel's method [16], and A7) unfilled sequences with holes. IRCCyN/IVC DIBR video datasets provide mean opinion scores (MOS) obtained by an Absolute Category Rating (ACR) subjective assessment experiment. A detailed description of the subjective assessment of IRCCyN/IVC DIBR video datasets can be found in [24].

3.2. Performance evaluation of the proposed objective quality assessment metric

To evaluate the performance of the proposed objective quality assessment metric, we used three performance measures: Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC), and root mean square error (RMSE).

In our experiment, we calculated the PLCC, SROCC, and RMSE between subjective MOS values provided in [24] and predicted MOS values transformed from objective quality assessments [25]. To obtain the predicted MOS values, objective quality assessment scores are transformed to the predicted MOS values (MOS_p) using the nonlinear regression with five parameters logistic function [25]. The predicted MOS values can be written as

$$MOS_p = \alpha_1 \left[\frac{1}{2} - \frac{1}{1+\exp(\alpha_2 (x - \alpha_3))} \right] + \alpha_4 x + \alpha_5 , \qquad (9)$$

where x represents the objective metric scores and the parameters (i.e., α_1 - α_5) are determined using the subjective MOS values and objective metric scores. To obtain the parameters, we used a curve fitting toolbox in MATLAB.

To verify the performance of the proposed method, seven existing quality assessments (QA) were used for performance comparisons. Five metrics were 2D image/video QA models. Two metrics were 3D image QA models for synthesized view. In 2D image QA models, PSNR, SSIM [23], multi-scale SSIM (MS_SSIM) [26], and visual information fidelity (VIF) [27] were used. Video quality measurement (VQM) is a 2D video QA model [28]. In 3D image QA models, view synthesis quality assessment (VSQA) [19] and 3DSwIM [20] for QA of synthesized view were used. Note that the 2D and 3D image QA models were applied to each frame. Then, the objective assessment scores at all frames were averaged to obtain final quality score.

Table 2. Prediction performance comparison for synthesized video.

Objective metrics	PLCC	SROCC	RMSE
PSNR	0.2605	0.2719	0.5545
SSIM	0.2685	0.2685	0.5652
MS-SSIM	0.4521	0.4529	0.4601
VIF	0.2768	0.2769	0.5628
VQM	0.3128	0.3143	0.5480
VSQA	0.5378	0.5381	0.4529
3DSwIM	0.5427	0.5441	0.4351
Proposed CTI	**0.7217**	**0.7218**	**0.4012**

Table 2 shows the prediction performance of the proposed and existing quality metrics. In Table 2, the performance evaluation results revealed the proposed CTI index achieved high correlation with subjective MOS of synthesized videos provided IRCCyN/IVC DIBR database (PLCC was 0.7217 and SROCC was 0.7218). In fact, the performance of the proposed objective QA metric was better than those of the existing 2D and 3D QA metrics. Naturally, these results indicate that the temporal inconsistency is one of the most important factors affecting the overall quality of the synthesized video. In particular, when SSIM applied to entire region at each frame of the synthesized videos, the performance prediction was poor (PLCC was 0.2685 and SROCC was 0.2685). On the other hand, the proposed method achieved high prediction performance by measuring the structural similarity only on the excessive flicker regions. These results point out that distortions on specific regions are highly related to the overall quality of synthesized video.

4. CONCLUSIONS

This paper presented a new objective temporal inconsistency measure, CTI, to effectively predict the quality of the synthesized video without reference video. The proposed CTI extracted excessive flicker regions. Then, the structural similarity on the excessive flicker regions between temporally neighboring frames was measured to quantify the perceptual effects of temporal inconsistency. Experimental results showed that the proposed method significantly improved the prediction performance of the synthesized video. In particular, the experimental results indicated the temporal distortions on specific regions were highly dependent on the overall quality of the synthesized video.

5. ACKNOWLEDGEMENT

This work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2015R1A2A2A01005724).

6. REFERENCES

[1] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. Kopylow, "A survey of 3DTV displays: Techniques and technologies," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1647-1658, Nov. 2007.

[2] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-dimensional displays: a review and applications analysis," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 362-371, Jun. 2011.

[3] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3DTV," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems*, 5291, pp. 93–104, May. 2004.

[4] P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3D video," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 453-465, Jun. 2011.

[5] O. Stankiewicz, K. Wegner, M. Tanimoto, and M. Domanski, "Enhanced View Synthesis Reference Software (VSRS) for Free-viewpoint Television," ISO/IEC JTC1/SC29/WG11 MPEG2013/M31520, 2013.

[6] C. Fehn, R. de la Barré, and S. Pastoor, "Interactive 3-D TV—concepts and key technologies," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 524–538, 2006.

[7] P. Ndjiki-Nya, M. Köppel, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T.Wiegand, "Depth image based rendering with advanced texture synthesis," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2010, pp. 424–429.

[8] H. G. Kim, Y. J. Jung, S. S. Yoon, and Y. M. Ro, "Multi-view stereo image synthesis using binocular symmetry based global optimization," in *Proc. SPIE Electronic Imaging, Stereoscopic Displays and Applications XXVI*, Mar. 2015.

[9] H. G. Kim, S. S. Yoon, and Y. M. Ro, "Temporally consistent hole filling method based on global optimization with label propagation for 3D video," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 3136-3140.

[10] E. Bosc, R. Pepion, P. Le Callet, M. Koppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, "Towards a new quality metric for 3-D synthesized view assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1332-1343, Nov. 2011.

[11] E. Bosc, P. Le. Callet, L. Morin and M. Pressigout, "Visual quality assessment of synthesized views in the context of 3DTV," *3D-TV system with Depth-Image-Based Rendering*, C. Zhu, Y. Zhao, L. Yu, and M. Tanimoto, Springer, New York, pp 439-473, 2013.

[12] M. Lambooij, W. A. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imaging Sci. Technol*., vol. 53, no. 3, pp. 1-14, 2009.

[13] H. G. Kim, Y. J. Jung, S. S. Yoon, and Y. M. Ro, "Investigating experienced quality factors in synthesized multi-view stereo images," in *Proc. Int. Conf. Digital Signal Processing*, Aug. 2014, pp. 568-573.

[14] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3-D warping using depth information for FTV," *Elsevier Signal Process.: Image Commun*., vol. 24, pp. 65–72, 2009.

[15] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View synthesis for advanced 3-D video systems," *EURASIP J. Image Video Process*., pp. 1-11, 2008.

[16] M. Köppel, P. Ndjiki-Nya, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand, "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1809–1812.

[17] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Using disparity for quality assessment of stereoscopic images," in *Proc. IEEE Int. Conf. Image Process.,* Oct. 2008, pp. 389-392.

[18] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. Int. Workshop Video Process. Qual. Metrics Consum. Electron*., Sep. 2010, pp. 4033-4036.

[19] P. H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," in *Proc. SPIE 8288*, USA, Jan. 2012.

[20] F. Battisti, E. Bosc, M. Carli, P. Le Callet, and S. Perugia "Objective Image Quality Assessment of 3D Synthesized Views," *Signal Process.: Image Commun.*, vol. 30, pp. 78-88, 2015.

[21] E. Bosc, M. K¨oppel, R. P´epion, M. Pressigout, L. Morin, P. Ndjiki-Nya, and P. Le Callet, "Can 3D synthesized views be reliably assessed through usual subjective and objective evaluation protocols?," in *Proc. IEEE Int. Conf. Image Process.*, 2011, pp. 2597–2600.

[22] A. Chamboller and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vision*, vol. 40, no. 1, pp. 120-145, 2011.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, 2004.

[24] E. Bosc, R. Pépion, P. Le. Callet, M. Köppel, P. Ndjiki-Nya, L. Morin, and M. Pressigout, "Perceived quality of DIBR-based synthesized views," in *Proc. of SPIE Optical Engineering + Applications,* San Diego, USA, Aug. 2011.

[25] H. Sheikh, M. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440-3451, 2006.

[26] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput*., Nov. 2003, pp. 1398–1402.

[27] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430-444, Feb. 2006.

[28] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast*., vol. 50, no. 3, pp. 312–313, Sep. 2004.

[29] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process*., vol. 19, no. 2, pp. 335–350, Feb. 2010.

[30] F. Zhang and D. Bull, "A perception-based hybrid model for video quality assessment," *IEEE Trans. Circuits Syst. Video Technol*., May 2015.

[31] P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most apparent distortion model for video quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2505–2508.

[32] D. Sandic-Stankovic, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological wavelets," *International Workshop on Quality of Multimedia Experience QoMEX*, May 2015.