Performance comparison of objective metrics on free-viewpoint videos with different depth coding algorithms

Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges

Performance comparison of objective metrics on free-viewpoint videos with different depth coding algorithms

Shishun Tian, Lu Zhang, Luce Morin, and Olivier Déforges

IETR UMR CNRS 6164, National Institute of Applied Sciences, Rennes, France

ABSTRACT

The popularity of 3D applications has brought out new challenges in the creation, compression and transmission of 3D content due to the large size of 3D data and the limitation of transmission. Several compression standards, such as, Multiview-HEVC and 3D-HEVC have been proposed to compress the 3D content aiding by view synthesis technologies, among which the most commonly used algorithm is Depth-Image-Based-Rendering (DIBR), but the quality assessment of DIBR-synthesized view is very challenging owing to its new types of distortions induced by inaccurate depth map which the conventional 2D quality metrics may fail to assess. In this paper, we test the performance of existing objective metrics on free-viewpoint video with different depth coding algorithms. Results show that all the existing objective metrics perform not well on this database including the full-reference and the no-reference. There is certainly room for further improvement for the algorithms.

Keywords: free-viewpoint video, DIBR, distortion, depth coding, quality assessment

1. INTRODUCTION

Providing more immersive perception of a visual scene, 3D video applications , such as 3D-TV[1] and Free-viewpoint TV (FTV),[2] have gained great public interest and curiosity in recent years. In most 3D applications, 3D content is obtained by using multiple cameras to record the same scene at slightly different viewpoints.

The Multiview-Video-Plus-Depth (MVD)[3] format is one of the most 3D representations, it consists of multiple texture views and their associated depth maps from different viewpoints. Furthermore, this 3D representation could be exploited by generating virtual viewpoints, thus the virtual view from a viewpoint which has not been recorded can be rendered by using Depth-Image-Based-Rendering (DIBR). MVD and DIBR can be used for many 3D applications, such as, Free-viewpoint TV (FTV) which is able to allow the users to view a 3D scene by freely changing their viewpoints. Efficient compression methods are also needed for 3D applications. Extensions of High-Efficiency-Video-Coding (HEVC): Multiview-HEVC (MV-HEVC) and 3D-HEVC have been developed by the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V). 3D-HEVC exploits additional coding tools and achieves the best compression performance for MVD data.[4]

Benefiting from efficient coding methods and DIBR, only limited number of original views need to be coded and transferred, the additional virtual views on the receiver side can be synthesized from the decoded views. However, this process will lead to some new kinds of distortions, which are very different from the distortions induced by 2D image compression. Since most video coding standards rely on Discrete Cosine Transform, which leads to specific distortions,[5] these distortions are often scattered over the whole image. Whereas the Depth-Image-Based-Rendering (DIBR) synthesized artifacts are mainly caused by depth compression and view synthesis, those usually happen in the dis-occluded areas. Especially, the inaccurate depth map may results in the displacement of the pixels in the synthesized virtual view, which causes geometric distortions. Therefore, the conventional 2D quality metrics, which focus on the common compression artifacts, may not be capable to evaluate DIBR view synthesis artifacts. Especially for free view-point videos, this problem becomes more challenging due to the lack of real reference.

The distortion in DIBR-synthesized views mainly originate from the pristine texture/depth image and the DIBR process as well. Several effort have been made to assess the quality of DIBR-synthesized views,[6–10] but

Further author information: (Send correspondence to Shishun Tian)
Shishun Tian: E-mail: shishun.tian@insa-rennes.fr

the influence of imperfect compressed depth map has been less discussed. Since the depth map is used to warp the pixel from the original viewpoint to the target viewpoint, the imperfect depth map may cause various structural geometric distortions in the synthesized views. In this paper, we specially focus on the impact of different compressed depth map on the synthesized view. The the performance of state-of-the-art objective quality metrics has been tested on free-viewpoint videos with different depth coding methods.

2. TESTED QUALITY METRICS

In this paper, we use 5 2D objective quality assessment metrics and 8 DIBR dedicated quality assessment metrics, including full-reference (FR), reduced-reference (RR) and no-reference (NR). Within which, there are two special metrics, they use the pristine image at the original viewpoint, from which the synthesized image at a new viewpoint is generated, as the reference image. They do not need the original image at the new viewpoint. That is not classical FR metric, so we call is side view based FR metric in this paper.

Tested 2D quality metrics:

- PSNR: the commonly used pixel-based Peak Signal-to-Noise-Ratio;
- SSIM: Structural SIMilarity, a widely used full-reference (FR) objective image quality assessment (IQA) metric proposed by Wang et al.[11]
- VIF: Visual Information Fidelity, the objective FR IQA metric proposed by Sheikh et al.[12] The image quality is calculated by quantifying the information loss between the reference and the distorted image;
- VIIDEO: Video Intrinsic Integrity and Distortion Evaluation Oracle, a completely blind objective video quality assessment (VQA) metric based on natural video statistics proposed by Mittal et al.[13] The visual video quality is obtained by measuring their deviations from these regularities.
- RRED: reduce-reference entropy-difference, a reduced-reference (RR) VQA metric proposed by Soundararajan et al.[14] They assume that the wavelet coefficients of natural image follow Gaussian scale mixture (GSM) distribution. The quality is obtained as the distance between the GSM distribution of the reference image and the distorted image.

Tested DIBR dedicated metrics:

- MW-PSNR: Multi-scale Wavelet PSNR, the MW-PSNR was the first FR metric applied on free viewpoint videos porposed by Sandić-Stanković.[15,16] It firstly decomposed the two views (reference and synthesized) using separable morphological wavelet transformation, then calculated the weighted sum of the mean squared errors (MSE) of all sub-bands at all the corresponding scales of the two views.
- MP-PSNR: Morphological pyramid PSNR, proposed by the same author[17] as MW-PSNR, MP-PSNR uses multi-scale decomposition as well. MP-PSNR uses morphological filters instead of morphological wavelet for decomposition. The geometric distortions are maintained by comparing the difference between the reference and synthesized image across different resolution levels.
- MW-PSNRr, MP-PSNRr: the reduced version proposed by the same authors,[18] which only use the detail images at higher decomposition levels.
- VQA-SIAT: an FR Video Quality Assessment metric proposed by Liu et al.[19] The quality score is obtained by measuring the temporal flickering and the distortion of spatio-temporal activity.
- NIQSV: No-reference Image Quality assessment of Synthesized-Views, a NR IQA metric proposed by Tian et al.[8] The distortion at each color component is captured by a pair of morphological erosion and dilation operations, then these distortions are integrated and weighted by an edge image since the DIBR synthesized artifacts mainly happen in around the object edges.

Tested DIBR dedicated side view based FR metrics:

- LOGS: the FR objective IQA metric proposed by Li et al.[20] based on measuring the LOcal Geometric distortions in the disoccluded region and global Sharpness.
- DSQM: DIBR-Synthesized image Quality Metric, proposed by Farid et al.[6] A block matching is firstly used to match the content in the reference image and the synthesized image, then the difference of Phase congruency (PC) in these two matched blocks is used to measure the quality of the block in the synthesized image. The final image quality is obtained by averaging the quality score of all the blocks.

3. EXPERIMENT

In this section, the performances of the above metrics are compared and analyzed. We firstly introduce the used database, and then present the performance comparison of these metrics.

3.1 Free-viewpoint Video database

The quality metrics are evaluated on the IRCCyN/IVC Free-viewpoint video (FVV) database[21] which is introduced in.[22] This FVV database contains six different multiview sequences: *Balloons* (1024 × 768), *BookArrival* (1024 × 768), *UndoDancer* (1920 × 1080), *GTFly* (1920 × 1080), *Kendo* (1024 × 768), *Newspaper* (1024 × 768). For each MVD sequence, two real views at one time instant t were chosen to generate 49 intermediate synthesized views in-between these two views using the View Synthesis Reference Software 1D Fast (VSRS-1D-Fast).[23] There are two different synthesis mode to be considered: *blended mode* and $non - blended\ mode$. The blended mode means the combination of the both view using a weighted blending based on the baseline distance, the non-blended mode means using the closer view for rendering and the other one for hole filling. A video sequence of 100 frames (at 10 fps) was built from the 49 intermediate virtual frames and 2 original real views to simulate a smooth camera motion from left to right and from right to left. as shown in Fig. 1. Where $O1$ and $O2$ denote the real views, $S1 \cdots S49$ denote the synthesized views.

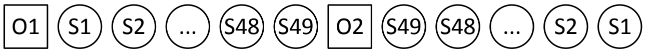


Figure 1. Video frame sequence

The associated depth maps were encoded through the following seven coding algorithms:

- C1: 3D-HEVC test model, 3D-HTM 0.4,[24] inter-view prediction and *View Synthesis Optimization* enabled;
- C2: Multiview Video Coding (MVC), JM 18.4;[25]
- C3: HEVC Test model, HM 6.1;[26]
- C4: JPEG2000, Kakadu implementation;[27]
- C5: based on,[28] a lossless-edge depth map coding based on optimize path and fast homogeneous diffusion;
- C6: based on,[29] this algorithm exploits the correlation with color frames;
- C7: Z-LAR-RP,[30] a region-based algorithm.

In this database, all coding algorithms were used in intra coding mode. Three quantization parameters were selected for each depth map codec test according to the visual quality of the rendered views. Views synthesized from compressed depth map and uncompressed depth map are used as the distorted and reference videos separately in this database in order to estimate the synthesis distortions produced by depth map compression.

3.2 Performance comparison

The reliability of objective metric could be evaluated by their correlation with subjective test scores. Differential Mean Opinion Score (DMOS) is usually used as subjective score. In this paper, the consistency of objective metrics was calculated by using Pearson Linear Correlation Coefficients (PLCC), Spearmans Rank Order Correlation Coefficients (SROCC) and Root-Mean-Square-Error (RMSE). Before measuring the PLCC, RMSE and SROCC values, the objective scores were fitted to the predicted DMOS ($DMOS_p$). The following nonlinear regression function should be used to map the objective scores to subjective scale, which is recommended by the Video Quality Expert Group (VQEG) Phase I FR-TV.[31]

$$DMOS_p = \beta_1(\frac{1}{2} - \frac{1}{1+e^{\beta_2(score-\beta_3)}}) + \beta_4 score + \beta_5 \tag{1}$$

where *score* is the score obtained by the objective metric and $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the parameters of this function. They are obtained through regression to minimize the difference between $DMOS_p$ and $DMOS$.

Table 1. Performance of each metrics on IVC Free Viewpoint Video database, where FR* indicates that the metric is side view based FR metric.

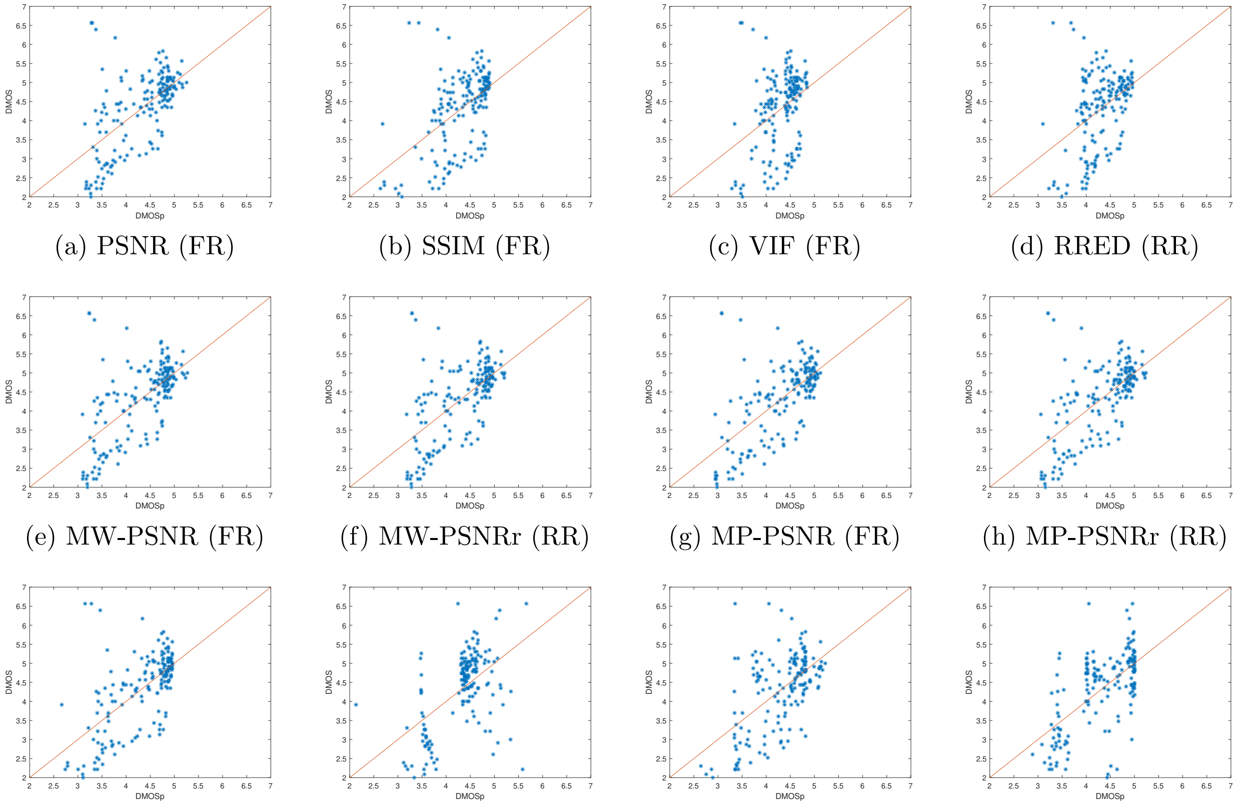
Metric		PLCC	RMSE	SROCC
2D metrics	PSNR (FR)	0.6052	0.7888	0.6043
	SSIM (FR)	0.5190	0.8471	0.5161
	VIF (FR)	0.3552	0.9264	0.3545
	VIIDEO (NR)	0.4011	0.9078	0.3631
	RRED (RR)	0.4129	0.9026	0.4035
DIBR metrics	MW-PSNR (FR)	0.6273	0.7708	0.6205
	MW-PSNRr (RR)	0.6028	0.7908	0.6066
	MP-PSNR (FR)	0.6466	0.7560	0.6355
	MP-PSNRr (RR)	0.6239	0.7745	0.6192
	VQA-SIAT (FR)	0.6287	0.7706	0.6232
	LOGS (FR*)	0.5398	0.8342	0.4292
	DSQM (FR*)	0.5766	0.8097	0.4938
	NIQSV (NR)	0.6192	0.7781	0.5508

For the image quality assessment metrics, the video quality scores are calculated by averaging the score of all the frames of each video. The obtained PLCC, RMSE, SROCC values are shown in Table 1. It can be observed that none of the tested quality metrics performs well on this Free Viewpoint Video database, since there is no metric whose PLCC is higher than 0.65. Generally, the DIBR metrics perform much better than conventional 2D metrics except PSNR, which is logical since these metrics are dedicated to handle the DIBR distortions.

The scatter plot of each quality metric is shown in Fig. 2. It seems that these methods have difficulties in predicting the worst or best qualities, which leave large empty regions in on the $DMOS_p$ axis compared to the $DMOS$ axis. Which is consistent with the results shown in Table 1 that no metric get a PLCC value bigger than 0.65. The metrics VIF, RRED and LOGS get the largest empty regions, they show lower correlation with the subjective results, which is similar to their correlation coefficients in Table 1. While most metrics succeed in assessing the videos with medium qualities.

The above methods compare the performance of each metric by calculating their correlations with the subjective results, however they just take the mean value of subjective scores into consideration, the uncertainty of the subjective scores has been ignored. In addition, the quality scores need to be regressed by a regression function cf. Eq. 1, that is not the way they are exactly used in real scenarios. Thus, we further conduct a statistical test proposed by Krasula et al.,[32] which does not suffer from the drawbacks of the above methods. The performance of objective metrics are evaluated by their classification abilities.

Firstly, the tested video pairs in the database are divided into two groups: different and similar. The first performance analysis is conducted by how well the objective metric can distinguish the video pair are significant different or not. If the two videos in the pair are significant different according to the subjective results, in the

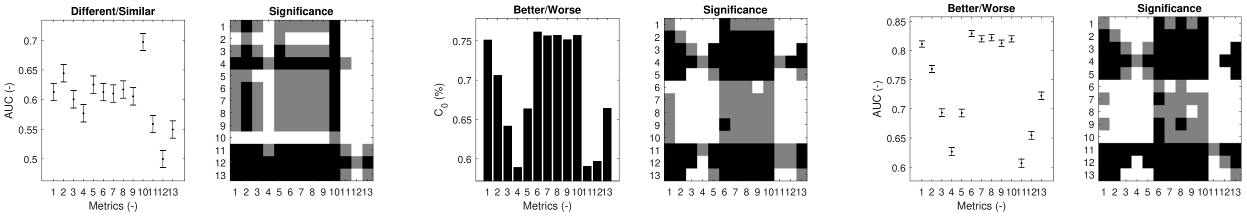


(a) PSNR (FR) (b) SSIM (FR) (c) VIF (FR) (d) RRED (RR)

(e) MW-PSNR (FR) (f) MW-PSNRr (RR) (g) MP-PSNR (FR) (h) MP-PSNRr (RR)

(i) VQA-SIAT (FR) (j) LOGS (FR*) (k) DSQM (FR*) (l) NIQSV (NR)

Figure 2. Scatter plots of DMOS versus DMOSp of each quality assessment method, where FR* indicates that the metric is side view based FR metric



(a) different / similar (AUC) (b) better / worse (C_0) (c) better / worse (AUC)

Figure 3. Results of Krasula's method, the metric numbers from 1 to 13 refer to: PSNR (1), SSIM (2), VIF (3), VIIDEO (4), RRED (5), MW-PSNR (6), MW-PSNRr (7), MP-PSNR (8), MP-PSNRr (9), VQA-SIAT (10), LOGS (11), DSQM (12) and NIQSV (13); the white color in the significant plots indicates that the performance in the row is significantly better and the metric in the column, similarly the block mean lower and gray means there is no significant difference between these two metrics.

second analysis, we determine whether the objective metric can correctly recognize the video of higher quality in the pair. The obtained results, the Area Under the Curves (AUC) and the Correct Classification percentage (C_0) are given in Fig. 3.

In the first different/similar analysis, the AUC values of most metrics are between 0.55 and 0.65, the VQA-SIAT performs significantly better than the others and the DSQM performs significant worse. In the better/worse analysis, the DIBR FR metrics perform significantly better than the 2D metrics except PSNR, but there is no AUC values bigger than 0.85, which is consistent with the results in Table 1 and Fig. 2 that none of the metrics can achieve a satisfactory correlation with the ground truth. Especially, the better/worse analysis results are consistent with the SROCC values, which give the monotonicity estimation, in Table 1. The metrics which lower SROCC values performs worse in the better/worse analysis according to Fig. 3 (b) and (c).

4. CONCLUSION

In this paper, we conducted a comparison of the performance of existing quality metrics on free-viewpoint videos with different depth coding algorithms. The results show that the DIBR dedicated metrics performs better than the conventional 2D quality metrics generally, but none of these metrics can achieve a satisfactory correlation with the ground truth since both the PLCC and SROCC values of these metrics are less than 0.65. There is certainly room for the improvement of quality assessment method of free-viewpoint videos with compressed depth maps.

REFERENCES

[1] Fehn, C., "A 3D-TV approach using depth-image-based rendering (DIBR)," in [*Proc. of VIIP*], **3**(3) (2003).

[2] Tanimoto, M., Tehrani, M. P., Fujii, T., and Yendo, T., "Free-viewpoint tv," *IEEE Signal Processing Magazine* **28**(1), 67–76 (2011).

[3] Merkle, P., Smolic, A., Muller, K., and Wiegand, T., "Multi-view video plus depth representation and coding," in [*2007 IEEE International Conference on Image Processing*], **1**, I – 201–I – 204 (Sept 2007).

[4] Mller, K., Schwarz, H., Marpe, D., Bartnik, C., Bosse, S., Brust, H., Hinz, T., Lakshman, H., Merkle, P., Rhee, F. H., Tech, G., Winken, M., and Wiegand, T., "3d high-efficiency video coding for multi-view video and depth data," *IEEE Transactions on Image Processing* **22**, 3366–3378 (Sept 2013).

[5] Yuen, M. and Wu, H., "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal processing* **70**(3), 247–278 (1998).

[6] Farid, M. S., Lucenteforte, M., and Grangetto, M., "Perceptual quality assessment of 3d synthesized images," in [*Multimedia and Expo (ICME), 2017 IEEE International Conference on*], 505–510, IEEE (2017).

[7] Farid, M. S., Lucenteforte, M., and Grangetto, M., "Evaluating virtual image quality using the side-views information fusion and depth maps," *Information Fusion* **43**, 47–56 (2018).

[8] Tian, S., Zhang, L., Morin, L., and Deforges, O., "NIQSV: A no reference image quality assessment metric for 3D synthesized views," in [*Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*], 1248–1252, IEEE (2017).

[9] Tian, S., Zhang, L., Morin, L., and Déforges, O., "NIQSV+: A No-Reference Synthesized View Quality Assessment Metric," *IEEE Transactions on Image Processing* **27**(4), 1652–1664 (2018).

[10] Tian, S., Zhang, L., Morin, L., and Deforges, O., "A full-reference image quality assessment metric for 3d synthesized views," *Electronic Imaging* **2018**(12), 366–1 (2018).

[11] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing* **13**(4), 600–612 (2004).

[12] Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," *IEEE Transactions on Image Processing* **15**, 430–444 (Feb 2006).

[13] Mittal, A., Saad, M. A., and Bovik, A. C., "A completely blind video integrity oracle," *IEEE Transactions on Image Processing* **25**(1), 289–300 (2016).

[14] Soundararajan, R. and Bovik, A. C., "Rred indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing* **21**(2), 517–526 (2012).

[15] Sandi-Stankovi, D., Battisti, F., Kukolj, D., Callet, P. L., and Carli, M., "Free viewpoint video quality assessment based on morphological multiscale metrics," in [*2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*], 1–6 (June 2016).

[16] Sandić-Stanković, D., Kukolj, D., and Le Callet, P., "DIBR synthesized image quality assessment based on morphological wavelets," in [*2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*], 1–6, IEEE (2015).

[17] Sandić-Stanković, D., Kukolj, D., and Le Callet, P., "Multi–scale synthesized view assessment based on morphological pyramids," *Journal of Electrical Engineering* **67**(1), 3–11 (2016).

[18] Sandić-Stanković, D., Kukolj, D., and Callet, P., "Dibr-synthesized image quality assessment based on morphological multi-scale approach," *EURASIP Journal on Image and Video Processing* **2017**(1), 4 (2016).

[19] Liu, X., Zhang, Y., Hu, S., Kwong, S., Kuo, C.-C. J., and Peng, Q., "Subjective and objective video quality assessment of 3d synthesized views with texture/depth compression distortion," *IEEE Transactions on Image Processing* **24**(12), 4847–4861 (2015).

[20] Li, L., Zhou, Y., Gu, K., Lin, W., and Wang, S., "Quality assessment of dibr-synthesized images by measuring local geometric distortions and global sharpness," *IEEE Transactions on Multimedia* **20**(4), 914–926 (2018).

[21] IVC-IRCCyN lab, "IRCCyN/IVC Free-Viewpoint synthesized videos database." http://ivc.univ-nantes.fr/en/databases/Free-Viewpoint_synthesized_videos/.

[22] Bosc, E., Hanhart, P., Le Callet, P., and Ebrahimi, T., "A quality assessment protocol for free-viewpoint video sequences synthesized from decompressed depth data," in [*2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*], 100–105, IEEE (2013).

[23] Zhang, L., Tech, G., Wegner, K., and Yea, S., "3d-hevc test model 5," *ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 JCT3V-E1005* (2013).

[24] "3D-HTM [online]." https://hevc.hhi.fraunhofer.de/.

[25] "JM [online]." http://iphome.hhi.de/suehring/tml/.

[26] "HM [online]." https://hevc.hhi.fraunhofer.de/.

[27] "Kakadu [online]." http://www.kakadusoftware.com/.

[28] Gautier, J., Meur, O. L., and Guillemot, C., "Efficient depth map compression based on lossless edge coding and diffusion," in [*2012 Picture Coding Symposium*], 81–84 (May 2012).

[29] Pasteau, F., Strauss, C., Babel, M., Déforges, O., and Bdat, L., "Adaptive color decorrelation for predictive image codecs," in [*2011 19th European Signal Processing Conference*], 1100–1104 (Aug 2011).

[30] Bosc, E., *Compression of Multi-View-plus-Depth (MVD) data: from perceived quality analysis to MVD coding tools designing*, PhD thesis, INSA de Rennes (2012).

[31] Group, V. Q. E., "Final report from the video quality experts group on the validation of objective models of multimedia quality assessment," *VQEG* (March 2008).

[32] Krasula, L., Fliegel, K., Le Callet, P., and Klíma, M., "On the accuracy of objective image and video quality models: New methodology for performance evaluation," in [*Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*], 1–6, IEEE (2016).