Signal, Image and Video Processing manuscript No.
(will be inserted by the editor)

No-Reference Quality Metric for HEVC Compression Distortion Estimation in Depth Maps

Muhammad Shahid Farid · Maurizio Lucenteforte · Marco Grangetto

Received: date / Accepted: date

Abstract Multiview video plus depth (MVD) is the most popular 3D video format due to its efficient compression and provision for novel view generation enabling the free-viewpoint applications. In addition to color images, MVD format provides depth maps which are exploited to generate intermediate virtual views using the depth image based rendering (DIBR) techniques. Compression affects the quality of the depth maps which in turn may introduce various structural and textural distortions in the DIBR synthesized images. Estimation of the compression related distortion in depth maps is very important for a high quality 3D experience. The task becomes challenging when the corresponding reference depth maps are unavailable e.g., when evaluating the quality on the decoder side. In this paper, we present a no-reference quality assessment algorithm to estimate the distortion in the depth maps induced by compression. The proposed algorithm exploits the depth saliency and local statistical characteristics of the depth maps to predict the compression distortion. The proposed 'depth distortion evaluator' (DDE) is evaluated on depth videos from standard MVD database compressed with the state-of-the-art high efficiency video coding (HEVC) at various quality levels. The results demonstrate that DDE can be used to effectively estimate the compression distortion in depth videos.

Keywords Gait recognition, Spatiotemporal features, Fisher vector encoding, Feature evaluation

M.S. Farid (✉)
Punjab University College of Information Technology, University of the Punjab, Lahore, Pakistan.
E-mail: shahid@pucit.edu.pk

M. Lucenteforte and M. Grangetto
Dipartimento di Informatica, Università degli Studi di Torino, Torino, Italy
E-mail: {lucente,grangetto@di.unito.it}

1 Introduction

Three dimensional (3D) video technology, such as 3D television (3DTV), 3D cinema, and free-viewpoint television (FTV), has received much popularity in the recent years. Compared to the traditional 2D video, the 3D videos produce more realistic display by providing the depth sensation to the viewer. Two views of a 3D scene captured at slightly different viewpoints are projected simultaneously, and a mechanism such as special active or passive eye-gears, is used to separate the view to respective eye to achieve the depth sensation. The recent 3D video applications e.g., free-viewpoint television (FTV) [1] and super multiview displays (SMV) [2] present the viewer a novel 3D view of the scene depending on his position.

Efficient representation of 3D content is momentous for success of 3D video applications. Various 3D content representations e.g., stereo video, multiview video, frame-compatible formats, and depth-based video formats have been proposed [3–8]. Multiview video plus depth (MVD) format [9] has been adopted by the MPEG for 3DTV and future FTV technologies [10] due to its efficient representation and compression, provision for intermediate virtual view generation and backward compatibility with the existing broadcasting infrastructure. In MVD format, in addition to color image, its per pixel depth values are also available which permits the generation of virtual views by using depth image based rendering (DIBR) technique [11].

In 3DTV and free-viewpoint applications, a large number of views (upto 64) are needed to provide a seamless horizontal parallax. However, capturing, coding and transmitting such a large number of views is not practical due to various cost, time, hardware, and bandwidth constraints. Therefore, only few camera views with corresponding depth maps are captured, coded and transmitted, and the rest of the views are generated on demand using DIBR.

Efficient compression of MVD data is a very important task in 3DTV processing chain to achieve a wide spread acceptability of 3DTV and FTV technologies. Recently, it has received significant research efforts resulting in new compression friendly MVD representations [12–17] and efficient 3D video encoders. Multi-view Video Coding (MVC) [18] was standardized as an extension of Advanced Video Coding (AVC) [19] for multiview video (MVV) format to support the stereoscopic and autostereoscopic 3D displays [20]. High Efficiency Video Coding (HEVC) [21] is the current state-of-the-art video encoder, which achieves the same subjective quality as AVC at approximately 50% less bit rate on average [22]. To extend HEVC to 3D video, Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) was jointly formed by the ITU-T Visual Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). The JCT-3V produced two extensions of HEVC: multiview video HEVC (MV-HEVC) [23,24] and 3D extension of HEVC (3D-HEVC) [25, 26]. The latter is the current state-of-the-art encoder for MVD data, which jointly encodes the texture and depth videos for increased coding efficiency. It achieves 50% and 20% bit rate saving over HEVC simulcast coding and MV-HEVC respectively [25].

Depth maps are texture-less, grayscale images which represent the geometry of the scene. While compression, the encoder may introduce various compression related artifacts in the depth images, which translate in various textural and structural distortions in the DIBR virtual image [27–31]. Thus the quality of the depth map is crucial for a faithful generation of the virtual views. Estimating the distortion in the depth maps induced by the compression is an important task to analyze the quality of the virtual image and end user 3D quality of experience (QoE). This task becomes more challenging when the corresponding reference depth videos are not available e.g., depth videos on the decoder side. Moreover, the recent researches on the quality assessment of DIBR synthesized images agree that in addition to texture images, the quality of the depth map must also be considered for a true judgment 3D quality [32,33]. In [34], a blind depth quality assessment metric (BDQM) is presented to predict the compression distortion in depth videos in absence of reference videos. The present research is based on BDQM, however it considers the visual attributes such as depth saliency and viewing conditions while evaluating the quality of depth videos. The main contributions of this research are:

- local statistical characteristics of depth image are analyzed to predict the compression distortion;
- depth map saliency is exploited to find the visually important regions which are weighted more than other regions while estimating the depth quality;
- the viewing conditions e.g., display resolution, viewer's distance from the display, affect the perceived quality of the color images as features at a certain scale may not reflect some distortions which are visible at other scales. A number of 2D image quality assessment (2D-IQA) algorithms e.g., [35]. take this factor into account and showed improved performance working in multi-scale fashion. This phenomenon has not been investigated in case of depth images. In this paper, we also design multi-scale versions of the proposed algorithm to see if it is helpful in quality assessment of depth images too.
- using the depth saliency and multi-scale phenomena, two quality metrics are designed and evaluated for performance;
- experiments are worked out on depth videos of standard MVD sequences compressed with HEVC encoder. The results demonstrate that the proposed DDE metric yields high correlation with the well-known full-reference 2D-IQA algorithms.

The rest of the paper is organized as follows: the related work is briefly presented in Section 2; the proposed DDE metric is described in Section 3, followed by experimental evaluation in Section 5. The conclusions are drawn in Sect. 6.

2 Related Work

Depth maps contain the geometry of a scene which is exploited in depth image based rendering (DIBR) to generate views for novel intermediate virtual viewpoints. Thus the quality of the depth maps directly affects the quality of the DIBR synthesized views, and hence the overall user experience. A number of algorithms have been proposed to evaluate the quality of synthesized and stereoscopic images. Many of them use the conventional 2D-IQA algorithms to estimate the quality of each texture image and combine the results. However, the researches, e.g., [32,33] tested various existing 2D-IQA algorithms and concluded that depth maps must also be considered for faithful quality assessment of 3D content.

Recently, a number of 3D image quality assessment (3D-IQA) algorithms have been proposed which use both the texture image quality and the depth map quality to evaluate the overall quality of 3D images [36]. 3D Vision based Quality Measure (3VQM) [37] estimates the ideal depth at each pixel value that could be used to obtain a distortion-free synthesized video. The ideal depth is compared with the given depth map to compute the spatial outliers, temporal outliers, and temporal inconsistencies, which are combined to estimate the quality of the DIBR synthesized image. The 3D-IQA algorithm proposed in [38] utilizes the two well-known 2D-IQA metrics: PSNR and SSIM [39] to assess the quality of 3D images. Using the depth images, a weight function is computed which controls the relative importance of

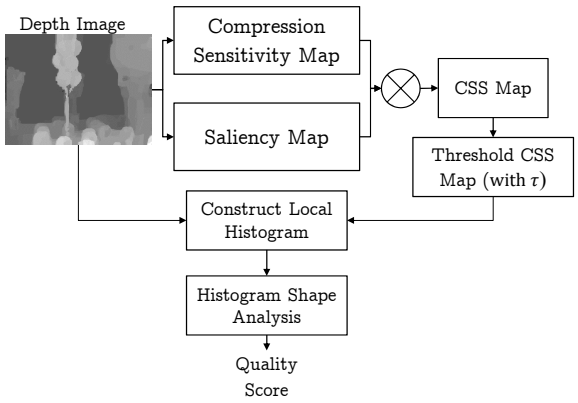


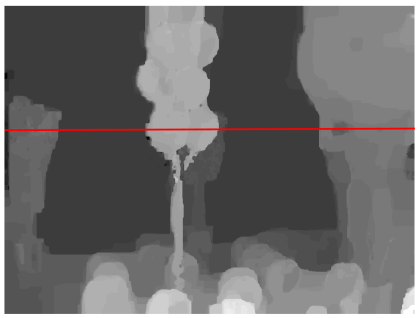
Fig. 1 Block diagram of the proposed single-stage depth distortion evaluator (DDE).

each pixel: the pixels closer to the camera belong to the foreground and hence are visually important. Such pixels are assigned more value compared to the far pixels. The value of PSNR or SSIM is then weighted using this weight function. The 3D-IQA proposed in [40] also uses the 2D-IQA algorithms (SSIM and C4) to estimate the texture image quality and compares the original and distorted depth images to estimate the depth quality. The two quality scores are then fused globally or locally to obtain the quality of the 3D images. The method in [41] proposed a modified SSIM [39] metric to estimate the film grain noise in HEVC compressed videos.

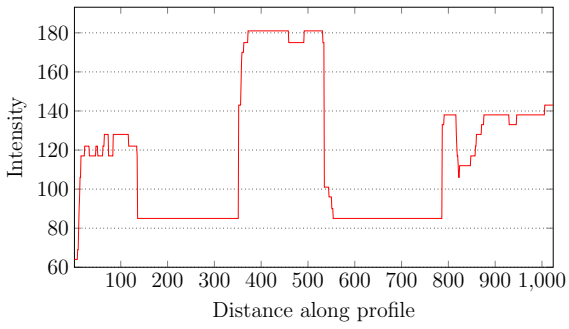
The algorithm presented in [42] estimates the impact of depth quality on the quality of the DIBR synthesized images. It estimates the distortion at each pixel by comparing the disparities of the reference and the distorted depth maps which is then integrated into a single score. All the 3D-IQA metrics described in this section target estimating the quality of the 3D images using the texture and depth images. Blind depth quality metric (BDQM) [34] is the only metric which is specifically designed to estimate the quality of the depth maps. It is a reference-less metric which constructs a local histogram at each compression sensitive pixel in the depth image and analyzes its shape to quantify the compression distortion. The proposed *depth distortion evaluator* (DDE) extends BDQM and exploits the depth visual saliency to find the compression sensitive locations in the depth image which are then used in quality evaluation. Moreover, the DDE algorithm is also proposed in multi-scale manner to assess the effect of viewing conditions on depth map quality.

3 Proposed Depth Distortion Evaluator (DDE)

Depth maps are texture-less, grayscale images which contain large homogeneous regions with sharp boundaries which correspond to the object edges. The values in the depth maps range from 0 to 255 and are inversely coded, that is closer



(a)



(b)

Fig. 2 (a) A sample depth image from MVD sequence Balloons, (b) intensity profile along red line segment in (a).

pixels have larger values. The large homogeneous regions in depth images are considered to be least affected by compression. On the contrary, the sharp discontinuities being difficult to encode are most affected by compression. The compression smoothens these sharp transitions and the proposed algorithm capitalizes this fact, it estimates the amount of smoothness introduced by compression and uses it to predict the compression distortion. The proposed algorithm works in four steps. In first step, the compression sensitivity map is computed; in second step the saliency of each depth pixel is computed. The two maps are geometrically integrated into a single map called distortion sensitivity map (DSM), which is representative of the compression sensitivity and saliency of the each depth pixel. The DSM map is thresholded to drop the less distortion sensitive pixels. In the third step, the BDQM [34] algorithm is used to estimate the distortion value for each distortion sensitive pixel. Finally, the distortion values are averaged to calculate the overall quality of the depth image. Fig. 1 shows the block diagram of the proposed single-stage DDE algorithm.

3.1 Computing depth compression sensitivity map

Depth images consist of linearly changing or homogeneous regions with sharp transitions. These regions usually represent the objects at different depth levels and the sharp transitions represent the edges or boundaries of these objects.

Fig. 2a shows a sample depth image where objects at different depth levels can be observed. Fig. 2b shows the intensity profile of a horizontal line shown in red in Fig. 2a. The intensity profile consists of constant line segments with sharp transitions, which correspond to the objects at different depth levels and their boundaries respectively. Compared to the homogeneous regions, these sharp boundary regions are highly susceptible to compression artifact which smoothens the sharp transitions. Fig. 3 shows the intensity profile of the pixels along the red horizontal line in Fig. 2a at different quality levels: uncompressed, compressed at QP 22, 38, and 46. It can be noted from these graphs that with the increase in the amount of compression the sharp transitions get smoother. Observe how the sharp transitions in original depth image (red plot) changed into smooth transitions at QP 46 (green plot) in Fig. 3. Therefore, the edge sharpness can be used to find the compression sensitive pixels in the depth image. Edge sharpness can be computed using gradient magnitude. We use Sobel operator to calculate the gradient of the depth map. Let I be a depth image, then the compression sensitivity map (CSM) is computed similar to [34] as:

$$CSM = \sqrt{G_x^2 + G_y^2} \tag{1}$$

where G_x and G_y are gradients along horizontal and vertical directions computed with Sobel filters:

$$G_x = I * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}; \; G_y = I * \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{2}$$

3.2 Computing depth visual saliency map

The importance of a depth pixel can be defined as the probability that a pixel is gazed at, and it is usually termed as depth saliency. Depth maps are very effective in predicting salient regions in 3D images [43] and a number of depth saliency techniques have been recently proposed, e.g., [44]. Generally, depth maps are exploited to find salient regions assuming that the objects closer to the camera get more visual attention. We exploit this notion here to find the visually important regions in depth images, and utilize this information in prioritizing the compression sensitive pixels in CSM (Eq. 1). The compression artifacts in highly salient regions would be more noticeable compared to the non-salient regions. Therefore, we propose to weight the CSM map with the depth visual saliency map (VSM) to find the depth locations which are visually important as well as are most susceptible to compression artifacts.

We use the saliency detection method proposed in [45] to estimate the VSM. It exploits the discrete cosine transform

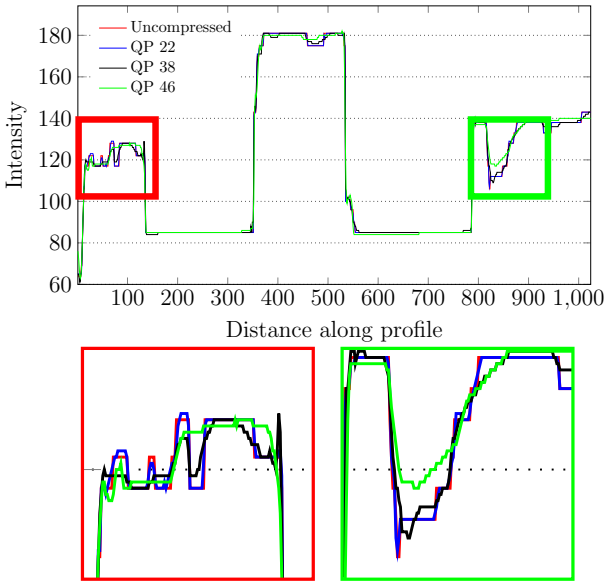


Fig. 3 Intensity profile of pixels along the horizontal line shown in red in Fig. 2a when the depth image is uncompressed, compressed at QP 22, 38, and 46.

(DCT) coefficients to extract depth features used to estimate the saliency. In particular, the depth image is divided into image patches of size 8×8. The DCT of each patch is computed and the DC coefficients are used as patch features. The saliency value of a patch is computed based on the contrast difference between this image patch and all the other image patches. The saliency value S_i of patch i is computed as:

$$S_i = \sum_{i \neq j} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{{d_{ij}}^2}{2\sigma^2}} U_{ij} \tag{3}$$

where d_{ij} represents the spatial distance between patches i and j, σ is the Gaussian model parameter which controls the degree of local and global contrast for the saliency estimation. The value σ is set to 5 as recommended in [45]. U_{ij} is the feature difference between the patches i and j, calculated as:

$$U_{ij} = \frac{|DC_i - DC_j|}{DC_i + DC_j} \tag{4}$$

where DC_i and DC_j are the DC coefficients of the image patches i and j respectively. The saliency map obtained from Eq. 3 is of smaller size ($\frac{M}{8} \times \frac{N}{8}$) and is up-sampled to original depth image size using bilinear interpolation.

3.3 Combining the CSM and VSM maps

The compression sensitivity map (CSM) and visual saliency map (VSM) are geometrically integrated to obtain a single map that we refer to as 'distortion sensitivity map' (DSM):

$$DSM = CSM^{\alpha} \cdot VSM^{\beta} \tag{5}$$

(a) (b) (c) (d) (e)

Fig. 4 (a) A sample depth image from MVD sequence Balloons, (b) compression sensitivity map, (c) visual saliency map, (d) distortion sensitivity map, (e) distortion sensitivity map after thresholding.

where α and β are control parameters used to adjust the relative importance of the two values. The values of α and β are empirically estimated and set to 0.5 and 0.3 respectively. Small DSM values correspond to the depth pixels which are not sensitive to compression artifacts, and thus they can be dropped from future processing to save the computation time. Only the depth pixels with DSM value greater than τ are considered in the next stage of the proposed algorithm. Fig. 4a shows a sample depth image of an MVD sequence. Fig. 4b and Fig. 4c show the CSM and VSM maps of 4a respectively. The combined DSM map is shown in Fig. 4d. By comparing the DSM with CSM, one can note that a number of pixels which were selected by CSM due sharp discontinuity are not that important in DSM due to their very low saliency values, e.g., pixels belonging to the contours of the background objects. The vice versa of this is also true. Fig. 4e shows the thresholded DSM when $\tau = 0.25$.

3.4 Computing the depth quality

In the final step of the proposed algorithm, we use the DSM map computed in the previous step to assess the quality of the depth image. For this purpose, we used blind depth quality metric (BDQM) [34]. BDQM constructs a histogram of the neighborhood of each DSM pixel and analyzes it to estimate the smoothness induced due to compression. Let I be a depth image whose quality is to be evaluated. Its DSM is computed as described earlier, and let S be the set of selected DSM pixels. For each $p_i \in S$, BDQM selects a patch P_i of size 15×15 centered at $I(x,y)$, where (x,y) is the location of p_i in DSM, and constructs the corresponding local histogram H_i^{κ} with κ equally sized bins. The quality index Q_i of p_i is defined as:

$$Q_i = \sum_{t=1}^{\kappa} [\max(H_i^{\kappa}) - H_i^{\kappa}(t)] \tag{6}$$

Finally, the Q_i value of all DSM pixels is averaged to obtain the quality of depth image I.

$$DDE = \frac{1}{|S|} \sum_{i=1}^{|S|} Q_i \tag{7}$$

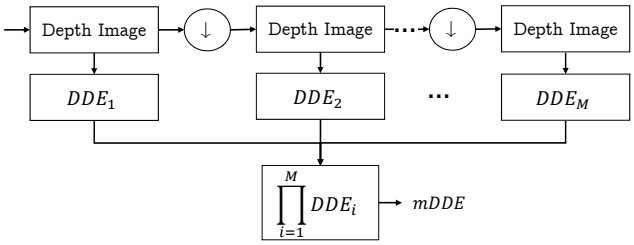


Fig. 5 Block diagram of multi-scale DDE. The symbol ↓ represents down-sampling by a factor of 2.

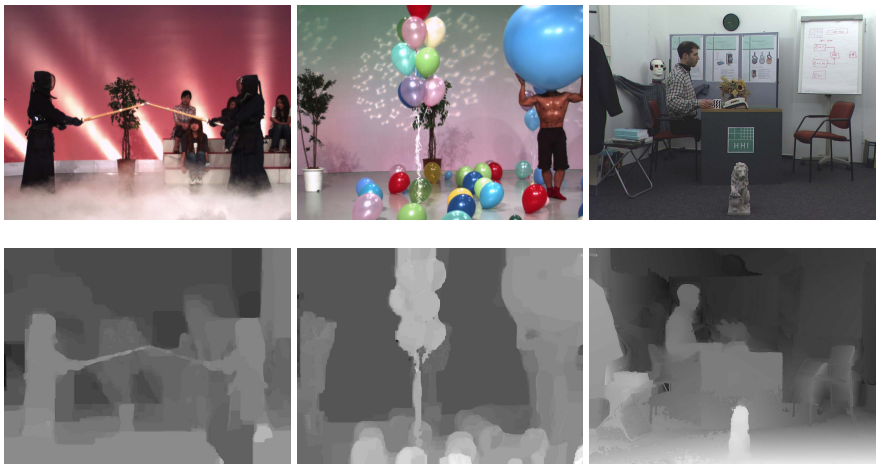
where $|S|$ represents the size of set S. DDE is a quality measure that is higher values represent better quality. We used the acronym DDE to differentiate it from BDQM in [34] as the former exploits the depth saliency too in evaluating the depth quality.

4 Multiscale Depth Distortion Evaluator

Viewing conditions, such as resolution of the display and viewer's distance from the display are considered important in evaluating the quality of color images. They affect the quality of the images as features at a certain scale may not reflect some distortions which are visible at other scales. A number of 2D image quality assessment (2D-IQA) algorithms e.g., [35]. consider this factor while evaluating the quality of images. These multi-scale 2D-IQA approaches showed better performance compared to their single-scale versions. However, this phenomenon has not been investigated in depth image quality assessment. In this section, we design multi-scale extension of BDQM and DDE algorithms to see if it improves their performance.

In multi-scale DDE algorithm (mDDE), we iteratively downsample the depth image by a factor of 2 and apply the DDE algorithm to estimate the quality of the resultant depth image at each scale. Let the number of scales be $1 \cdots M$ with scale 1 representing the original image. The quality scores obtained at each scale are combined to obtain the final quality score:

$$mDDE = \prod_{i=1}^{M} DDE_i^{w_i} \tag{8}$$



(a) (b) (c)

Fig. 6 Sample texture and depth images of the left view of 3D video sequences from the test dataset, (a) Kendo, (b) Balloons, and (c) Book Arrival.

Table 1 Test database details: number of frames in the video (#F), selected views (V) and frame rate (FR).

Sequence	#F	V	View Size	FR	Provider
Kendo	300	{1,5}	1024 × 768	30	Nagoya Univ.
Balloons	300	{1,5}	1024 × 768	30	Nagoya Univ.
Book Arrival	100	{10,8}	1024 × 768	16	Fraunhofer HHI

where w_i is used to control the relative contribution of score at the ith scale. Fig. 5 shows the block diagram of multi-scale DDE algorithm. Multi-scale BDQM algorithm (mBDQM) is designed analogously.

5 Experimental Evaluation and Results

The performance of the proposed depth map quality assessment algorithms: DDE and mDDE is evaluated on standard MVD depth videos compressed with state-of-the-art HEVC encoder [21]. We also compared the performance of the proposed metrics with BDQM [34] and its multi-scale version mBDQM. The test database is built from three standard MVD sequences: Kendo, Balloons and Book Arrival. Tab. 1 presents the details of the selected sequences. Two views of each sequence are selected and the corresponding depth videos are compressed with HEVC[1] at 10 different quality levels, namely QP=$\{10,14,18,22,26,30,34,38,42,46\}$. Thus 20 full length depth videos were obtained for each sequence, and a total of 60 videos were generated. Sample frames from the test dataset are shown in Fig. 6.

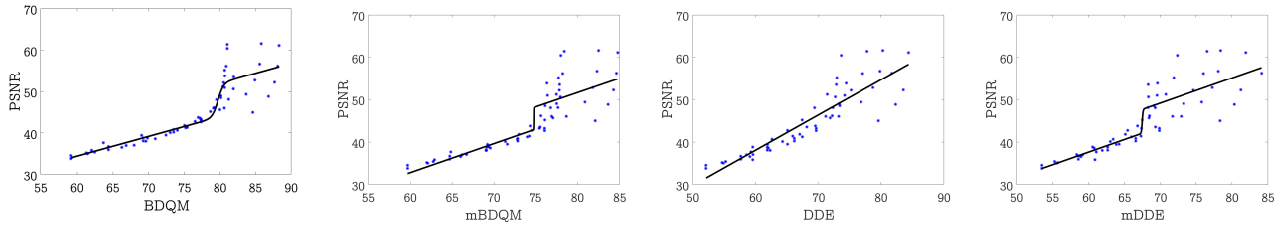
We compare the proposed quality assessment algorithms with Peak Signal to Noise Ratio (PSNR) to evaluate their performance. Since depth maps are texture-less, gray-scale images therefore the visual 2D-IQA algorithms are not effective to measure their quality. PSNR is usually used for this purpose. The performance of the four depth quality assessment algorithms: BDQM, mBDQM, DDE, and mDDE is evaluated using Pearson linear correlation coefficient (PLCC) for *prediction accuracy* test, and Root Mean Square Error (RMSE) is used to estimate the *prediction error*. Before computing PLCC and RMSE measures, according to the Video Quality Expert Group (VQEG) recommendations [46]. the algorithm predicted scores Q are mapped to PSNR with a monotonic nonlinear regression function. The regression mapping is performed using the following logistic function outlined in [47]:

$$Q_p = \beta_1 \left(\frac{1}{2} - \frac{1}{\exp(\beta_2 (Q - \beta_3))} \right) + \beta_4 Q + \beta_5 \tag{9}$$

where Q_p are the mapped score and $\beta_1, \cdots, \beta_5$ are the regression model parameters.

In all experiments, the BDQM parameters are set to the default values. The value of number of scales M (Eq. 8) depends on the viewing distance as well as the resolution of the image. For both mDDE and mBDQM algorithms, we tested different values of M and found that $M = 2$ achieves the best results. The values of w_1 and w_2 for respective scales are also computed empirically and they are set to 0.6 and 0.4 respectively. Tab. 2 compares the performance of the proposed algorithms with BDQM and mBDQM in terms of PLCC on each test sequence. The results show that the proposed DDE algorithm outperforms the other three quality metrics in all experiments. It achieves a high correlation with full reference PSNR metric in each test sequence, with an average PLCC of 0.9445. Tab. 3 shows that the DDE metric has lowest prediction error in all experiments. Both the prediction

[1] Version HM 11.0 of the HEVC reference software with Main profile.



(a) (b) (c) (d)

Fig. 7 Scatter plots between PSNR and mapped scores for the proposed and the compared 3D-IQA algorithms: (a) BDQM, (b) mBDQM, (c) DDE, (d) mDDE.

Table 2 Performance Comparison in terms of Pearson Linear Correlation Coefficient (PLCC)

Sequence	BDQM	mBDQM	DDE	mDDE
Kendo	0.9027	0.8901	0.9147	0.9040
Balloons	0.8453	0.8375	0.9355	0.9076
Book Arrival	0.9227	0.8869	0.9832	0.8902
Overall	0.8902	0.8715	0.9445	0.9006

Table 3 Performance Comparison in terms of Root Mean Square Error (RMSE)

Sequence	BDQM	mBDQM	DDE	mDDE
Kendo	3.6697	3.8876	3.4463	3.6467
Balloons	4.2190	4.3145	2.7901	3.3144
Book Arrival	2.2630	2.7124	1.0714	2.6741
Overall	3.3839	3.6382	2.4359	3.2117

accuracy and prediction error results show that DDE metric performs better than other single and multi-scale depth quality assessment algorithms.

Scatter plot is another effective mean to evaluate the efficiency of image quality assessment metrics. It is used to analyze the correlation between the mapped predicted scores and the subjective rankings. In Fig. 7, the scatter plots between PSNR and mapped scores for the proposed and the compared 3D-IQA algorithm are presented. Compared to the other methods, the scatter plot of the proposed metric shows better fitting between the estimated and the reference scores which indicates its ability to better estimate the quality of the compressed depth maps.

It can be observed from the results presented in Tab. 2 and Tab. 3 that the multi-scale versions of both BDQM and DDE algorithms performed rather poor compared to their single-scale versions. This might be due to the texture-less nature of the depth images, which means that in contrast to the color images the quality of the depth images do not vary with the change in the viewing distance. This assertion, however, requires more investigation through subjective testings which is in our future work plan. A software release of the proposed DDE metric is available at[2].

6 Conclusions

Depth maps are important in multiview video plus depth (MVD) format where they are exploited to generate images at novel viewpoints using depth image based rendering techniques. Depth maps are grayscale texture-less images which provide per pixel depth value of the corresponding color image. The compression may introduce various artifacts in depth maps, which in turn result in structural and textural distortion in the DIBR images. In this paper, we presented a no-reference quality assessment algorithm to estimate the HEVC compression artifacts in depth maps. The proposed algorithm uses depth visual saliency and local statistical characteristics of the depth maps to estimate the compression distortion. The experiments performed on standard depth video sequences demonstrate the effective of the proposed algorithm.

[2] http://www.di.unito.it/~farid/3DQA/DDE.html

References

1. M. Tanimoto. FTV: Free-viewpoint Television. *Signal Process.-Image Commun.*, 27(6):555 – 570, 2012.
2. M.P. Tehrani et al. Introduction of super multiview video systems for requirement discussion. document ISO/IEC JTC1/SC29/WG11/M31052, Geneva, CH, Oct 2013.
3. A. Vetro, A. M. Tourapis, K. Muller, and T. Chen. 3D-TV content storage and transmission. *IEEE Trans. Broadcast.*, 57(2):384–394, June 2011.
4. K. Muller, P. Merkle, G. Tech, and T. Wiegand. 3D video formats and coding methods. In *Proc. Int. Conf. Image Process. (ICIP)*, pages 2389–2392, 2010.
5. A. Smolic, K. Mueller, P. Merkle, P. Kauff, and T. Wiegand. An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution. In *Proc. Picture Coding Symp. (PCS)*, pages 1–4, 2009.
6. G. Ballocca, P. D'Amato, M. Grangetto, and M. Lucenteforte. Tile format: A novel frame compatible approach for 3d video broad-

casting. In *Proc. IEEE Int. Conf. Multimed. and Expo (ICME)*, pages 1–4, July 2011.
7. A. Vetro. Frame compatible formats for 3D video distribution. In *Proc. Int. Conf. Image Process. (ICIP)*, pages 2405–2408, Sept 2010.
8. Marco Cagnazzo, Béatrice Pesquet-Popescu, and Frédéric Dufaux. *3D Video Representation and Formats*, pages 102–120. John Wiley & Sons, Ltd, 2013.
9. A. Smolic et al. Multi-view video plus depth (MVD) format for advanced 3D video systems. doc. ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q6, 2007.
10. Call for proposals on 3D video coding technology, Mar 2011. documnet ISO/IEC JTC1/SC29/WG11.
11. C. Fehn. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XI*, volume 5291, pages 93–104, 2004.
12. M. Domanski et al. High efficiency 3D video coding using new tools based on view synthesis. *IEEE Trans. Image Process.*, 22(9):3517–3527, 2013.
13. M.S. Farid, M. Lucenteforte, and M. Grangetto. Panorama view with spatiotemporal occlusion compensation for 3D video coding. *IEEE Trans. Image Process.*, 24(1):205–219, Jan 2015.
14. T. Maugey, A. Ortega, and P. Frossard. Graph-based representation for multiview image geometry. *IEEE Trans. Image Process.*, 24(5):1573–1586, May 2015.
15. B.O. Ozkalayci and A.A. Alatan. 3D planar representation of stereo depth images for 3DTV applications. *IEEE Trans. Image Process.*, 23(12):5222–5232, Dec 2014.
16. M. S. Farid, M. Lucenteforte, and M. Grangetto. A panoramic 3d video coding with directional depth aided inpainting. In *Proc. Int. Conf. Image Process. (ICIP)*, pages 3233–3237, Oct 2014.
17. A. I. Purica, E. G. Mora, B. Pesquet-Popescu, M. Cagnazzo, and B. Ionescu. Multiview plus depth video coding with temporal prediction view synthesis. *IEEE Trans. Circuits Syst. Video Technol.*, 26(2):360–374, Feb 2016.
18. P. Merkle, A. Smolic, K. Muller, and T. Wiegand. Efficient prediction structures for multiview video coding. *IEEE Trans. Circuits Syst. Video Technol.*, 17(11):1461–1473, Nov 2007.
19. T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Technol.*, 13(7):560–576, July 2003.
20. Jinmi Kang and Kidong Chung. High-performance depth map coding for 3D-AVC. *Signal, Image and Video Processing*, 10(6):1017–1024, Sep 2016.
21. G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand. Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1649–1668, Dec 2012.
22. J. Ohm, G.J. Sullivan, H. Schwarz, Thiow Keng Tan, and T. Wiegand. Comparison of the coding efficiency of video coding standards – including high efficiency video coding (HEVC). *IEEE Trans. Circuits Syst. Video Technol.*, 22(12):1669–1684, Dec 2012.
23. G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro, and Y. K. Wang. Overview of the multiview and 3d extensions of high efficiency video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(1):35–49, Jan 2016.
24. G.J. Sullivan et al. Standardized Extensions of High Efficiency Video Coding (HEVC). *IEEE J. Sel. Topics Signal Process.*, 7(6):1001–1016, Dec 2013.
25. K. Muller et al. 3D High-Efficiency Video Coding for Multi-View Video and Depth Data. *IEEE Trans. Image Process.*, 22(9):3366–3378, Sept 2013.
26. Ruihai Jing, Qian Zhang, Bin Wang, PengTao Cui, Tao Yan, and Jifeng Huang. Cart-based fast cu size decision and mode decision algorithm for 3D-HEVC. *Signal, Image and Video Processing*, 13(2):209–216, Mar 2019.
27. P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Müller, P.H.N. de With, and T. Wiegand. The effects of multiview depth video compression on multiview rendering. *Signal Process.-Image Commun.*, 24(1–2):73 – 88, 2009.
28. L. Fang, N. M. Cheung, D. Tian, A. Vetro, H. Sun, and O. C. Au. An analytical model for synthesis distortion estimation in 3D video. *IEEE Trans. Image Process.*, 23(1):185–199, Jan 2014.
29. M.S. Farid, M. Lucenteforte, and M. Grangetto. Edge enhancement of depth based rendered images. In *Proc. Int. Conf. Image Process. (ICIP)*, pages 5452–5456, Oct 2014.
30. G. Leon, H. Kalva, and B. Furht. 3D Video Quality Evaluation with Depth Quality Variations. In *Proc. 3DTV Conf., True Vis.,-Capture Transmiss. Display 3D Video (3DTV-CON)*, pages 301–304, May 2008.
31. M.S. Farid, M. Lucenteforte, and M. Grangetto. Edges shape enforcement for visual enhancement of depth image based rendering. In *Proc. Int. Workshop Multimed. Signal Process. (MMSP)*, pages 406–411, 2013.
32. 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 J. Sel. Topics Signal Process.*, 5(7):1332–1343, Nov 2011.
33. 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 for Consum. Electron. (VPQM)*, 2010.
34. M. S. Farid, M. Lucenteforte, and M. Grangetto. Blind depth quality assessment using histogram shape analysis. In *Proc. 3DTV Conf., True Vis.,-Capture Transmiss. Display 3D Video (3DTV-CON)*, pages 1–5, July 2015.
35. H. Zhang, Y. Huang, X. Chen, and D. Deng. MLSIM: A multi-level similarity index for image quality assessment. *Signal Process.-Image Commun.*, 28(10):1464 – 1477, 2013.
36. Payman Aflaki, Miska M. Hannuksela, and Moncef Gabbouj. Subjective quality assessment of asymmetric stereoscopic 3D video. *Signal, Image and Video Processing*, 9(2):331–345, Feb 2015.
37. M. Solh, G. AlRegib, and J.M. Bauza. 3VQM: A vision-based quality measure for DIBR-based 3D videos. In *Proc. IEEE Int. Conf. Multimed. and Expo (ICME)*, pages 1–6, July 2011.
38. E. Ekmekcioglu, S. Worrall, D. De Silva, A. Fernando, and A. M. Kondoz. *Depth Based Perceptual Quality Assessment for Synthesised Camera Viewpoints*, pages 76–83. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
39. Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, April 2004.
40. A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau. Quality assessment of stereoscopic images. *EURASIP J. Image Video Process.*, 2008, 2009.
41. Seongwan Kim, Daehyun Pak, and Sangyoun Lee. Ssim-based distortion metric for film grain noise in HEVC. *Signal, Image and Video Processing*, 12(3):489–496, Mar 2018.
42. W. D. Jang, T. Y. Chung, J. Y. Sim, and C. S. Kim. FDQM: Fast quality metric for depth maps without view synthesis. *IEEE Trans. Circuits Syst. Video Technol.*, 25(7):1099–1112, July 2015.
43. C. Koch and S. Ullman. Shifts in selective visual attention: Towards the underlying neural circuitry. In LuciaM. Vaina, editor, *Matters of Intelligence*, volume 188 of *Synthese Library*, pages 115–141. Springer Netherlands, 1987.
44. J. Wang, M.P. DaSilva, P. LeCallet, and V. Ricordel. Computational model of stereoscopic 3d visual saliency. *IEEE Trans. Image Process.*, 22(6):2151–2165, June 2013.
45. Y. Fang, J. Wang, M. Narwaria, P. Le Callet, and W. Lin. Saliency detection for stereoscopic images. *IEEE Trans. Image Process.*, 23(6):2625–2636, June 2014.
46. VQEG. RRNR-TV Group Test Plan, 2007. Version 2.2.
47. H.R. Sheikh, M.F. Sabir, and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.*, 15(11):3440–3451, Nov 2006.