BLIND DEPTH QUALITY ASSESSMENT USING HISTOGRAM SHAPE ANALYSIS

Muhammad Shahid Farid, Maurizio Lucenteforte, Marco Grangetto

Dipartimento di Informatica, Università degli Studi di Torino
Corso Svizzera 185, 10149 Torino, Italy

ABSTRACT

Multiview videos plus depth (MVD) is a popular 3D video representation where pixel depth information is exploited to generate additional views to provide 3D experience. Quality assessment of MVD data is of paramount importance since the latest research results show that existing 2D quality metrics are not suitable for MVD. This paper focuses on depth quality assessment and presents a novel algorithm to estimate the distortion in depth videos induced by compression. The proposed algorithm is no-reference and does not require any prior training or modeling. The proposed method is based solely on the statistical analysis of the compression sensitive pixels of depth images. The experimental results worked out on a standard MVD dataset show that the proposed algorithm exhibits a very high correlation with conventional full-reference metrics.

Index Terms — Depth image quality metric, Free-viewpoint TV, Depth image based rendering, Quality assessment

1. INTRODUCTION

Multiview-video-plus-depth format for 3D content representation has been adopted for current and future 3D television technologies e.g. free-viewpoint television (FTV) [1] and Super Multiview (SMV) displays [2]. The gray scale depth image represents the per pixel depth value of the corresponding texture image which is exploited to generate novel views through depth image based rendering (DIBR) [3]. In MVD format only few views with their associated depth maps are coded and transmitted.

The compression of MVD data is indeed an important activity in 3D television framework and much attention has been devoted to this research area. To efficiently compress MVD data various coding formats have been proposed and new tools have been developed, e.g., [4–7]. Advanced Video Coding (H.264/AVC) [8] has been used in past to encode the texture videos and depth videos independently, also known as simulcast coding. The novel High Efficiency Video Coding (HEVC) [9] is the current state of the art video coding tool. The Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) has recently developed extensions of HEVC to efficiently encode multiview videos and MVD data. Multiview-HEVC (MV-HEVC) [10] extends the HEVC syntax to encode MVD without additional coding tools whereas 3D-HEVC [11] is expressively dedicated to the design of novel coding techniques for MVD. 3D-HEVC encodes the base view with its depth map using unmodified HEVC whereas the dependent views and their depth maps are encoded by exploiting additional coding tools. 3D-HEVC achieves the best compression ratio for MVD data [11]. To achieve autostereoscopy additional intermediate viewpoints can be generated with DIBR on the receiver side. Given a DIBR algorithm, the perceptual quality of the rendered images depends on both texture and depth image quality. Quality of the depth map is particularly important as the compression artifacts in depth maps can cause structural and textural distortions in the synthesized image [12–14] resulting in poor 3D experience.

3D image and video quality assessment is a more difficult and complex problem compared to its 2D counterpart. Earlier, 2D image quality metrics have been used to quantify the quality of 3D images (video plus depth) and stereoscopic images. In this context, the 2D metrics have been used in two ways: some metrics estimate the quality by assessing each texture image separately and aggregating the values without considering the depth images. Others exploit the depth maps in addition to texture image quality to predict the overall quality. However, due to different nature of acquisition, representation, transmission and rendering, 3D images are affected by different types of quality artifacts [15, 16]. Recent studies [17, 18] tested various existing 2D image quality metrics to assess the quality of stereoscopic and 3D images and concluded that none of the existing 2D quality metrics is suitable in this context.

Ekmekcioglu et.al [19] proposed a 3D quality assessment algorithm based on weighted PSNR and SSIM [20]. They propose to weight each pixel quality value (PSNR or SSIM) with the corresponding depth value to increase the contribution of pixels closer to the camera; indeed, according to their study the closer the pixel, the larger the impact on visual perception. The 3D QA proposed in [21] combines SSIM and C4 [22] with disparity estimation to compute a single quality metric. The two measures are then integrated (globally or locally) to obtain the final quality value. Boev et. al. [23] proposed a full-reference multi-scale stereo video QA algorithm that computes the monoscopic artifacts from the texture images and stereoscopic artifacts from the disparity images. Cyclopean images are constructed from the reference and the test stereopairs with block based matching; SSIM is used to quantify the monoscopic artifacts (2D artifacts like blur, noise, etc). The perceptual disparity maps computed for test and reference stereopairs are compared to estimate the binocular distortions (e.g. keystone, color distortion).

Most existing 3D quality metrics are full reference and few consider depth maps in the evaluation. As already described, quality of depth images is very important due to their role in intermediate view generation. Moreover, no-reference 3D quality evaluation is of fundamental importance since the corresponding original views may not be available; indeed, cost, hardware and bandwidth constraints usually impose to capture a limited set of views and the quality of the synthesized view must be estimated in absence of the corresponding reference. Furthermore, as the depth im-

This work was partially supported by Universitá degli Studi di Torino and Compagnia di San Paolo under project AMALFI (ORTO119W8J).

ages are gray scale textureless images usually consisting of large homogeneous or linearly changing regions with sharp edges representing objects' boundaries, the conventional 2D visual quality metrics such as SSIM [20] are not effective to assess the quality of depth images. As an answer to the mentioned issues, this paper proposes 'Blind depth quality metric' (BDQM), a no-reference algorithm to assess the quality of compressed depth images. The major contributions of the paper are:

- the proposal of a novel no-reference depth quality metric BDQM for blind evaluation of depth compression artifact;
- the shape of the histogram of compression sensitive depth pixels is exploited to estimate the depth quality; in particular, we show that as the compression ratio is increased the histogram around depth transitions flattens because of smoothing;
- BDQM is used to predict the quality of depth images undergoing HEVC compression at various bitrates.

The rest of the paper is organized as follows. In Sect. 2 the proposed algorithm is described. In Sect. 3 experimental results and comparisons with existing techniques are presented. The research is concluded in Sect. 4 with a discussion on its various aspects and possible applications as future work.

2. PROPOSED DEPTH IMAGE QUALITY METRIC

The proposed quality metric works in two steps; first, the compression sensitivity map (CSM) of the depth image is computed to locate the pixels which are the most susceptible to compression artifacts. Second, for each compression sensitive pixel (CSP) a histogram of the neighborhood is constructed and analyzed to determine the quality index. BDQM builds on the key observation that the histogram around a CSP gets flattened when increasing the amount of compression; indeed, compression mostly affects the sharp discontinuities of the depth image. The proposed algorithm exploits the shape of the histogram to predict depth quality. The following subsections describe each step in detail.

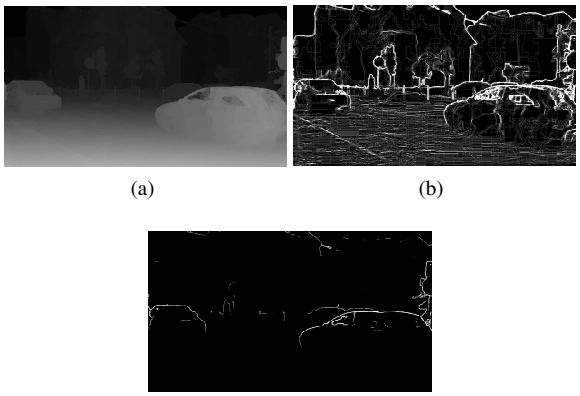
2.1. Computing compression sensitivity map

It is well known that the boundary regions between objects at different depth levels are highly susceptible to compression artifact compared to the flat homogeneous areas of depth images. Therefore, the magnitude of the depth gradient can be a simple and effective means for evaluating compression sensitivity. Let I be an $M \times N$ depth image. The compression sensitivity map (CSM) of I is computed from its gradient magnitudes as:

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

where G_x and G_y are gradients along horizontal and vertical directions and can be computed with Sobel filters.

The gradient magnitude can be used to select the compression sensitive depth pixels that will be used to estimate the quality index in the following section. Fig. 1a shows a depth image from Poznan_Street sequence (View 5, 1st Frame) and its corresponding gradient representing the CSM (Fig. 1b). Thresholding by dropping the pixels with $CSM \leq \tau$ is used to locate the most compression sensitive pixels; please note that this choice has also a positive side effect since it dramatically reduces the computational cost of the whole metric. As an example, Fig. 1c shows the CSM after thresholding with $\tau = 4$.



(a) (b)

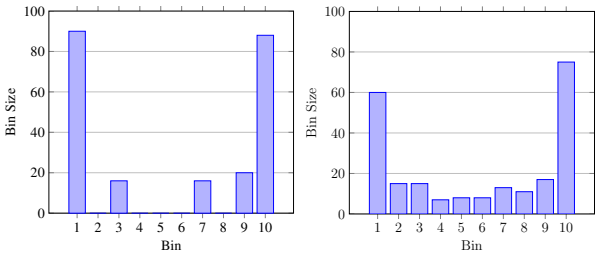
(c)

Figure 1. Depth saliency detection: a depth image (a), its CSM (b), thresholded CSM ($\tau = 4$) (c).

2.2. Depth quality index

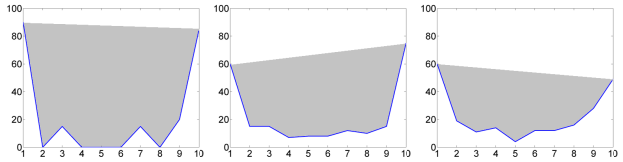
The CSM computed in the previous step is used to estimate the quality of the depth image. The CSPs defined above belong to the sharp discontinuities representing the boundaries between two usually very flat or linearly changing regions at different depth levels. To quantify the effect of compression the neighborhood of the CSPs is examined to determine the smoothness induced by quantization. A local histogram is constructed and analyzed to infer the presence of compression effects. As the CSPs lie on or in the proximity of the boundary between two different depth levels, the histogram appears to be very peaked around two bins. In presence of compression, the depth transitions tend to be smoothed and the effect can be captured by a local histogram where the two peaks are less pronounced and the values are more equally distributed in between. Fig. 2 shows a sample histogram of a CSP (neighborhood of size 15×15) from Poznan_Street sequence compressed with HEVC with quantization parameter QP=30 (Fig. 2a) and QP=42 (Fig. 2b), respectively. The histogram is computed onto 10 equal bins. Two very high peaks with values above 85 can observed in Fig. 2a showing that the depth values are concentrated around two bins whereas the rest of the histogram is very sparse and almost empty. In Fig. 2b it can be noted that the histogram of the same region exhibits lower peaks and higher valley between them when QP=42: a drop of 30 and 15 can be observed in the two peaks respectively along with increased values of the bins in-between. We can conclude that higher compression makes the histogram flatter.

To predict the quality of a depth image we propose to estimate the histogram dispersion by measuring the area which lies on top



(a) (b)

Figure 2. Histogram of a salient pixel from Poznan_Street test sequence, view 5, frame 1 at QP=30 (a), and QP=42 (b).



(a) (b) (c)

Figure 3. Predicting the quality index. (a) QP=30, $\mathcal{Q}_i$ = 675, (b) QP=42, $\mathcal{Q}_i$ = 525, (c) QP=46, $\mathcal{Q}_i$ = 375.

of the histogram curve (see gray area in Fig. 3): the larger the area, the less compressed is depth. An area value is associated to each CSP and then averaged together to compute the final quality index. Let $\mathcal{S}$ be the set of CSPs of depth image I and let $p_i \in \mathcal{S}$ be a CSP with coordinates $(x, y) \mid \{1 \leq x \leq M; 1 \leq y \leq N\}$. For each $p_i \in \mathcal{S}$, we select a patch $\mathcal{P}_i$ of size $w \times w$ centered at (x, y) and calculate the corresponding local histogram. Let $\mathcal{H}_i^{\kappa}$ denote the histogram distribution of patch $\mathcal{P}_i$ with κ equally sized bins. The quality index $\mathcal{Q}_i$ of p_i is defined as:

$$\mathcal{Q}_i = \sum_{t=1}^{\kappa} [\max(\mathcal{H}_i^{\kappa}) - \mathcal{H}_i^{\kappa}(t)] \tag{2}$$

Fig. 3 graphically shows the proposed quality index. The figure shows distribution curves of a sample CSP of the first frame of Poznan_Street test sequence coded by HEVC with different QP. The blue line represents the histogram distribution whereas the area inside the curve is shadowed in gray. One can note that, as we conjectured above, the histogram area is decreasing when QP increases. Finally, the $\mathcal{Q}_i$ value of all CSPs is averaged to obtain the quality of depth image I.

$$BDQM = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} \mathcal{Q}_i \tag{3}$$

where $|\mathcal{S}|$ represents the size of $\mathcal{S}$. Blind depth quality metric (BDQM) is computed for each frame of the depth video and the values are averaged to predict the quality of a whole video sequence. BDQM is a quality measure that means the larger the value of BDQM is, the better the quality of the depth map.

3. EXPERIMENTAL EVALUATION

In this section the proposed BDQM is tested on a number of standard depth videos undergoing HEVC compression. Each depth video sequence is encoded at 6 different compression levels, namely QP={26,30,34,38,42,46} using version HM 11.0 of the HEVC reference software with Main profile. We selected HEVC as a benchmark for depth coding since the most promising future 3D video coding standards, e.g. 3D-HECV, will leverage on it. The goal of our analysis here is to show that the no reference BDQM can compete with full reference metrics. Since depth maps are textureless gray-scale images the visual image quality metrics are not effective to assess their quality. Peak Signal to Noise Ratio (PSNR) is usually used to evaluate the quality of depth maps. We compare BDQM with PSNR to evaluate its performance. In the following we employ 5 depth videos from standard sequences in the MPEG and HHI datasets (see details in Tab. 1). The coded depth quality is evaluated using the proposed BDQM with parameters $w = 15$, $\tau = 5$ and $\kappa = 10$ and compared with the PSNR computed versus the uncoded reference.

To evaluate the performance of BDQM we chose Pearson linear correlation coefficient (PLCC) for *prediction accuracy* test and Spearman rank order correlation coefficient (SROCC) and Kendall rank order correlation coefficient (KROCC) for *Prediction Monotonicity* test. To estimate the *prediction error* we compute Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) measures. Before computing these performance parameters, according to Video Quality Expert Group (VQEG) recommendations [24] the BDQM predicted scores Q are mapped to PSNR with a monotonic nonlinear regression function. The following logistic function outlined in [25] is used for regression mapping:

$$Q_p = \beta_1 \left(\frac{1}{2} - \frac{1}{\exp \beta_2 (Q - \beta_3)} \right) + \beta_4 Q + \beta_5 \tag{4}$$

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

The performance parameters discussed above are reported in Tab. 2 for each test sequence. The table shows that the proposed BDQM achieves very high correlation with PSNR in every experiment with an average PLCC of 0.9920. The SROCC and KROCC are equal to 1 in all experiments as the predicted scores are monotonic. The average prediction error in terms of RMSE and MAE turns to be 0.29 and 0.25, respectively. All the collected results demonstrate the accuracy of the proposed quality metric. To further evaluate the reliability of BDQM the performance parameters have been computed over the entire dataset, i.e. without considering the 5 videos as separated experiments; such an approach allows one to understand if BDQM can be used not only to rank the quality of different compression levels of the same content but also to compare different scores of different videos. The results of this global analysis are shown in Tab. 3. The PLCC achieved over the entire dataset turns to be 0.9076, showing again high correlation between BDQM and PSNR. The values of SROCC and KROCC are equal to 0.8439 and 0.7089 respectively, demonstrating the good monotonicity between the two metrics also when BDQM is used to compare different video contents. Clearly, the statistics presented in Tab. 2 and 3 show that the quality scores predicted by the proposed metric are quite accurate and reliable. Finally, in Fig. 4 we show the scatter plot of the predicted scores versus PSNR over the complete dataset to let the reader visually appreciate the obtained level of correlation. More details on experimental evaluation and a software release of the proposed BDQM metric can be found at: http://www.di.unito.it/~farid/3DQA/BDQM.html.

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

Sequence	#F	V	View Size	FR	Provider
Poznan_Hall2	200	7	1920 × 1088	25	Poznan Univ. of Tech.
Poznan_Street	250	5	1920 × 1088	25	Poznan Univ. of Tech.
Kendo	300	1	1024 × 768	30	Nagoya University
Balloons	300	1	1024 × 768	30	Nagoya University
Book_Arrival	100	10	1024 × 768	16	Fraunhofer HHI

Table 2. Performance Evaluation of proposed BDQM.

Sequence	PLCC	RMSE	MAE
Poznan_Hall2	0.9808	0.6056	0.5131
Poznan_Street	0.9941	0.2438	0.2036
Kendo	0.9985	0.1588	0.1276
Balloons	0.9978	0.1554	0.1466
Book_Arrival	0.9889	0.3187	0.2796
Average:	0.9920	0.2965	0.2541

Table 3. Performance of BDQM over entire dataset.

PLCC	SROCC	KROCC	RMSE	MAE
0.9076	0.8439	0.7089	1.7498	1.4902

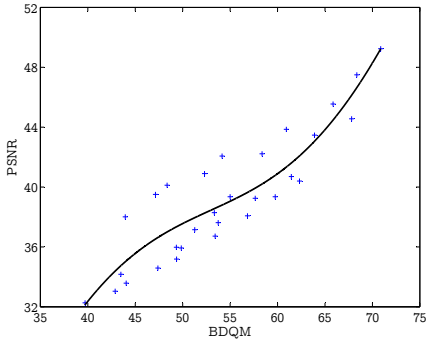


Figure 4. Scatter plot of BDQM versus PSNR over entire dataset and curve fitted with logistic function.

4. CONCLUSIONS AND FUTURE WORK

In this paper a novel no-reference metric able to rank the compression artifacts of depth maps has been presented. The proposed algorithm leverages on the observation that depth images are characterized by flat regions with sharp boundaries that are potentially blurred after compression. The proposed algorithm estimates depth quality by measuring the blurriness of the compression sensitive regions of the depth image using a histogram based approach. The experimental results show that BDQM exhibits high prediction accuracy when compared to full reference PSNR metric.

BDQM can be integrated with no-reference image quality metrics to design novel 3D image quality scores that, in addition to texture image also consider the depth image to better estimate the overall quality. Another future application that we foresee is the use of BDQM within the rate distortion optimization stage of depth map compression algorithms. Since BDQM is based on the estimation of the quality of sharp transitions in the depth map it is expected to be a valuable instrument for predicting textural and structural distortions in synthesized images.

5. REFERENCES

[1] M. Tanimoto, "FTV: Free-viewpoint Television," *Signal Process.-Image Commun.*, vol. 27, no. 6, pp. 555 – 570, 2012.

[2] M.P. Tehrani et al., "Proposal to consider a new work item and its use case - rei : An ultra-multiview 3D display," *ISO/IEC JTC1/SC29/WG11/m30022*, July-Aug 2013.

[3] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *SPIE Electron. Imaging*, 2004, pp. 93–104.

[4] M. Domanski et al., "High efficiency 3D video coding using new tools based on view synthesis," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3517–3527, 2013.

[5] M.S. Farid et al., "Panorama view with spatiotemporal occlusion compensation for 3D video coding," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 205–219, Jan 2015.

[6] T. Maugey, A. Ortega, and P. Frossard, "Graph-based representation for multiview image geometry," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1573–1586, May 2015.

[7] M.S. Farid et al., "A panoramic 3D video coding with directional depth aided inpainting," in *Proc. Int. Conf. Image Process. (ICIP)*, Oct 2014, pp. 3233–3237.

[8] T. Wiegand et al., "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, July 2003.

[9] G.J. Sullivan et al., "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.

[10] G.J. Sullivan et al., "Standardized Extensions of High Efficiency Video Coding (HEVC)," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1001–1016, Dec 2013.

[11] K. Muller et al., "3D High-Efficiency Video Coding for Multi-View Video and Depth Data," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3366–3378, Sept 2013.

[12] P. Merkle et al., "The effects of multiview depth video compression on multiview rendering," *Signal Processing: Image Communication*, vol. 24, no. 1, pp. 73–88, 2009.

[13] M.S. Farid, M. Lucenteforte, and M. Grangetto, "Edges shape enforcement for visual enhancement of depth image based rendering," in *IEEE 15th Int. Workshop Multimedia Signal Process. (MMSP)*, 2013, pp. 406–411.

[14] M.S. Farid, M. Lucenteforte, and M. Grangetto, "Edge enhancement of depth based rendered images," in *Proc. Int. Conf. Image Process. (ICIP)*, 2014, pp. 5452 – 5456.

[15] Q. Huynh-Thu, P. Le Callet, and M. Barkowsky, "Video quality assessment: From 2d to 3d - challenges and future trends," in *Proc. ICIP*, Sept 2010, pp. 4025–4028.

[16] F. Speranza et al., "Effect of disparity and motion on visual comfort of stereoscopic images," in *SPIE Electron. Imaging*, 2006.

[17] E. Bosc et al., "Towards a new quality metric for 3-d synthesized view assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1332–1343, Nov 2011.

[18] P. Hanhart and T. Ebrahimi, "Quality assessment of a stereo pair formed from decoded and synthesized views using objective metrics," in *Proc. 3DTV-CON*, Oct 2012, pp. 1–4.

[19] E. Ekmekcioglu et al., "Depth based perceptual quality assessment for synthesised camera viewpoints," in *User Centric Media*, vol. 60, pp. 76–83. 2012.

[20] Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, April 2004.

[21] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, 2009.

[22] M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *Proc. ICIP*, Sept 2003, vol. 3, pp. 2284–2298.

[23] A. Boev et al., "Towards compound stereo-video quality metric: a specific encoder-based framework," in *IEEE Southwest Symp. Image Anal. Interp.*, 2006, pp. 218–222.

[24] VQEG, "RRNR-TV Group Test Plan," 2007, Version 2.2.

[25] 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.*, vol. 15, no. 11, pp. 3440–3451, Nov 2006.