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

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com

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

Muhammad Shahid Farid*, [a, b], Maurizio Lucenteforte[a], Marco Grangetto[a]

[a] Dipartimento di Informatica, Università degli Studi di Torino, Torino, Italy
[b] College of Information Technology, University of the Punjab, Lahore, Pakistan

ARTICLE INFO

Keywords:
3D image quality assessment
Information fusion
View synthesis
Depth image based rendering
Free-viewpoint TV

ABSTRACT

Three Dimensional (3D) image quality assessment is a challenging problem as compared to 2D images due to their different nature of acquisition, representation, coding, and display. The additional dimension of depth in multiview video plus depth (MVD) format is exploited to obtain images at novel intermediate viewpoints using depth image based rendering (DIBR) techniques, enabling 3D television and free-viewpoint television (FTV) applications. Depth maps introduce various quality artifacts in the DIBR-synthesized (virtual) images. In this paper, we propose a novel methodology to evaluate the quality of synthesized views in absence of the corresponding original reference views. It computes the statistical characteristics of the side views from whom the virtual view is generated, and fuses this information to estimate the statistical characteristics of the cyclopean image which are compared to those of the synthesized image to evaluate its quality. In addition to texture images, the proposed algorithm also considers the depth maps in evaluating the quality of the synthesized images. The algorithm blends two quality metrics, one estimating the texture distortion in the synthesized texture image induced by compression, transmission, 3D warping, or other causes and the second one determining the distortion of the depth maps. The two metrics are combined to obtain an overall quality assessment of the synthesized image. The proposed Synthesized Image Quality Metric (SIQM) is tested on the challenging MCL-3D and SIAT-3D datasets. The evaluation results show that the proposed metric significantly improves over state-of-the-art 3D image quality assessment algorithms.

1. Introduction

Image quality assessment (IQA) has been a major research area since the last three decades due to its vast applicability in modern broadcasting systems and due to the increasing demand of high picture quality by the end user. In the recent years, the advancements in 3D television, cinema and other advanced displays, e.g., free-viewpoint television (FTV) [1] and super multiview (SMV) displays [2], have posed new challenges in terms of quality assessment of 3D pictures. 3D image and video quality assessment is a more difficult and complex problem compared to its 2D counterpart. Due to different nature of acquisition, representation, transmission, and rendering of 3D images, they suffer from different types of quality artifacts [3–6]. Moreover, the additional dimension of depth maps in 3D content also introduces various quality artifacts. Research studies [7–9] suggest that in addition to texture image quality, depth map quality must also be incorporated in the evaluation process to assess the true quality of 3D images.

The latest autostereoscopic and multiview autostereoscopic displays, e.g., FTV, SMV allow the user to enjoy the 3D scene by interactively controlling the viewpoint. Such technologies require a huge number of views to provide a smooth motion parallax. However, capturing, coding, and transmitting such a large number of views is not practical due to various cost, hardware, and bandwidth constraints. Therefore, limited camera views are captured and transmitted and the additional intermediate viewpoints are generated with depth image based rendering (DIBR) techniques [10]. Given a DIBR algorithm, the perceptual quality of the rendered views depends on both the texture image quality and the depth map quality [7,8]. The quality of the synthesized views is important as in a multiview autostereoscopic environment, most of the views presented to the viewer are virtually gener-

* Corresponding author.
Email addresses: shahid@pucit.edu.pk, farid@di.unito.it (M.S. Farid); lucente@di.unito.it (M. Lucenteforte); grangetto@di.unito.it (M. Grangetto)
URLs: http://www.di.unito.it%7Efarid/; http://www.di.unito.it/%7Emgrange/

https://doi.org/10.1016/j.inffus.2017.11.007
Received 25 February 2017; Received in revised form 5 November 2017; Accepted 12 November 2017
Available online xxx

ated. The quality of these views, thus, has significant impact on the overall user experience. Moreover, being able to predict the quality of virtual views can be exploited also in compression algorithms to drive the rate distortion optimization stage. As an example, the VSO coding tool of 3D-HEVC [11] exploits such prediction to guarantee the quality of the virtual views.

The quality of a synthesized image is affected by a number of 2D and 3D artifacts that depend on many factors, such as:

- depth maps are noisy and imperfect since these are generally estimated through stereo-matching algorithms [12], and may cause structural and textural distortion in the synthesized images [13–15]. Moreover, compression of texture images and depth maps introduces additional artifacts [3,16];
- large discontinuities in the depth maps introduce holes in the synthesized images. Holes also appear as a consequence of re-sampling when warped pixel locations need to be mapped to integer coordinates. These holes are filled through inpainting approaches which may introduce blur in the final picture;
- other 3D rendering and viewing artifacts, e.g., binocular rivalry, visual discomfort and ghosting or crosstalk [17–20] have a significant impact on the perceived quality.

These issues make the quality assessment (QA) of 3D synthesized images more complex compared to 2D-IQA. Moreover, in absence of the original reference images, the QA of virtual images becomes more challenging problem. As already recalled, for virtual viewpoints the corresponding ground truth video is generally missing. In Fig. 1, a typical scenario for the quality assessment of a virtual view is shown. An intermediate virtual view V'_k, whose corresponding reference view V_k is not available, is generated from the distorted left and right views V_i and V_j through DIBR. In absence of the reference kth view, common full-reference image quality metrics cannot be employed: an algorithm capable to predict the quality of the synthesized image exploiting only the available side views is needed to evaluate the quality of the synthesized view V'_k.

2. Related work

Various stereoscopic image quality assessment algorithms have been proposed in literature [21–26], however the quality assessment of DIBR-synthesized images is relatively less investigated. To quantify the structural distortion in synthesized view due to DIBR, Bosc et al. [27] compared the edges of the original and the warped images. However, this metric is limited to structural distortion estimation and cannot be used to represent the overall quality of the virtual image as it does not compute the color related artifacts. CSED (Color and Sharpness of Edge Distortion) [28] is another full reference quality metric that targets the hole regions to assess the color distortion and uses the edge sharpness of the reference and virtual images to assess structural distortion. The algorithm in [29] compares the regions of high spatial frequency of the stereopair images to estimate the contrast and luminance changes.

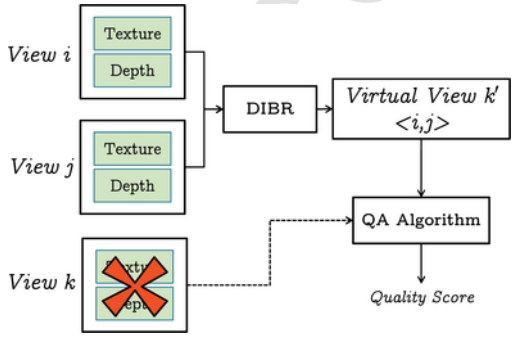

Fig. 1. A typical DIBR-Synthesized view quality assessment scenario: the quality of intermediate virtual view V'_k is subject to evaluation whose corresponding reference view V_k is not available.

Kim et al. [30] proposed to apply a weighting map to the conventional 2D quality metrics such as SSIM [31] and peak signal to noise ratio (PSNR). The weighting map is computed by combining the depth maps and the motion information estimated from the texture images. You et al. [32] proposed to use the disparity information to improve the performance of 2D quality metrics on 3D data. The 3D-IQA algorithm in [33] estimates the structural distortion in the synthesized image using Hausdorff distance and combines it with SSIM score. 3DSwIM [34] metric detects the human skin regions in the virtual image and finds the corresponding regions in the reference images; the two corresponding regions are compared to determine the quality of the virtual picture.

De Silva et al. [25] proposed a learning based Stereoscopic Structural Distortion (StSD) metric to evaluate the quality of 3D videos. The quality assessment algorithm proposed in [35] estimates the quality of the synthesized image by measuring how classes of image contours change due to synthesis process. The no-reference synthesized image quality metric proposed in [36] exploits simple morphological operators to predict the quality of the virtual image. They use the opening and closing morphological operations to remove the synthesis distortions in the virtual image. This filter image is then compared with the synthesized image to estimate its quality. The 3D quality estimator proposed in [37] combines SSIM and C4 [38] and also exploits the disparity to estimate the quality. View Synthesis Quality Assessment (VSQA) [39] combines SSIM with three weighting functions derived from contrast, orientation and texture maps of the reference and synthesized views to assess the quality of virtual pictures. A good literature on 3D-IQA and various 3D quality artifacts can be found in [40,41].

Most existing quality assessment algorithms for DIBR synthesized images are full-reference and rely on the conventional 2D-IQA algorithms. As described earlier, in modern 3DTV and FTV applications few sparse viewpoints are captured and DIBR is exploited to obtain a large number of intermediate views to support smooth horizontal parallax. The existing 3D-IQA metrics cannot be used to evaluate the quality of these novel views as the corresponding ground truth videos are not available. In this paper, we propose a new methodology to assess the quality of virtual views in absence of the corresponding references. We propose to use the original texture and depth images from whom the virtual image is generated as references to estimate the quality of the virtual image. This concept is novel and has not been explored in the past for 3D image quality assessment. Based on this concept, we present a novel 3D-IQA algorithm to estimate the quality of the synthesized images in absence of corresponding reference images. The major contributions of this paper are as follows:

- The overall contribution of this paper is the proposal of a novel 3D-IQA algorithm to evaluate the quality of virtual images obtained with depth image based rendering (DIBR) techniques. The proposed metric considers not only the texture images but also the corresponding depth maps in assessing the quality of the synthesized image. The two quality scores are combined to obtain the overall synthesized image quality.
- A novel Texture Distortion Metric (TDM) is proposed to assess the quality of the synthesized texture image exploiting information fusion theory. The novelty of TDM lies in the use of cyclopean eye theory and divisive normalization transform (DNT) in context of DIBR synthesized image quality assessment. In particular, the histograms

of the DN transformed texture input images are fused together to estimate the statistical characteristics of the mental (cyclopean) image. The histogram of the DNT of synthesized image is also computed and compared with that of the mental image to estimate the quality of the synthesized texture image. Such approach to texture quality evaluation has been previously introduced in [42], where it has been analyzed in presence of HEVC compression only using objective metrics.

- In addition to texture images, depth maps are also considered in quality evaluation; indeed, distortion in depth maps significantly affects the sharp boundaries in the depth map which in turn causes structural distortion in the synthesized image. Based on this observation, a novel Depth Distortion Metric (DDM) is presented. The DDM locates the noise sensitive pixels in the original depth maps and use them to evaluate the quality of the distorted depth maps using histogram shape analysis technique. A similar approach for totally blind evaluation of depth maps quality has been recently proposed in [43] in the particular case of HEVC depth compression only.
- It must be noted that the proposed 3D-IQA algorithm estimates the quality of the synthesized image without using the reference image corresponding to the synthesized viewpoint. It instead uses the left and the right view images and depth maps, which have been used to synthesize the virtual image, to evaluate the synthesized image quality. This is also a novelty of the proposed algorithm, as the existing 3D-IQA algorithms, except the no-reference techniques, use the reference images corresponding to the virtual viewpoint to assess the quality of the synthesized images.

The proposed SIQM metric is compared with the mean opinion score (MOS) of subjective tests reported by 3D synthesized image datasets, MCL-3D [44] and SIAT-3D [45]) targeting different kind of distortions ranging from transmission loss errors to JPEG compression. The obtained results show the excellent correlation of SIQM with subjective rankings compared with state-of-the-art 3D-IQA algorithms.

The rest of the paper is organized as follows: Section 3 presents an overview of the proposed quality metric and describes the texture distortion metric (TDM). Section 4 describes the depth distortion metric (DDM) and the combination of TDM and DDM to get a single global metric. The experimental evaluation is carried out in Section 5 and the conclusions are drawn in Section 6.

3. Proposed synthesized image quality metric

The proposed SIQM targets the quality assessment of a virtual image that is generated from two distorted stereograms and their corresponding depth maps by using DIBR. In particular, SIQM is a combination of *Texture Distortion Metric* (TDM) and *Depth Distortion Metric* (DDM). The TDM exploits the divisive normalization transform and cyclopean eye theory to estimate the quality of the synthesized texture image. Since the corresponding reference image is not available, TDM estimates a cyclopean image from the reference left and right texture images. It estimates the statistical characteristics of the input views and fuses them together to estimate the statistical characteristics of the cyclopean image, that is then compared to the synthesized image to predict the distortion. The DDM estimates the distortion in the depth images that are used in the synthesis process. Fig. 2 shows the block diagram of SIQM. In the diagram V_l, V_r and D_l, D_r, are the left and the right distorted stereoscopic images with corresponding depth images respectively which are used to generate the intermediate virtual image V_s by DIBR. The goal is to estimate the quality of V_s in absence of the corresponding original view. On the other hand the reference textures and depths V_l', V_r', D_l' and D_r' are available and can be exploited in the evaluation. In the following we assume that all images have resolution $M \times N$.

3.1. Divisive normalization transform and cyclopean perception models

Our TDM quality model is based upon the Cyclopean Perception theory [46] proposed by Béla Julesz in 1971. The Cyclopean Perception analyzes stereopsis and refers to the formation of a virtual image in our mind (Julesz called it *'the cyclopean retina for stereopsis'*) from stimuli received from the left and the right eye. The cyclopean image, also referred to as *mental image*, is a view obtained by fusing the left and the right views as if it was captured by a virtual eye (usually referred to as the cyclopean eye) placed in between the two eyes.

Divisive normalization (DN) model is based upon the standard psychophysical and physiological models that have been used to study the nonlinear behaviors of cortical neuron in biological vision [47]. The use of DN model in image quality assessment was pioneered by Teo and Heeger in [48]. It uses a set of wavelet-like linear sensors to obtain a scale and orientation decomposition of the image. The output of these sensors is weighted and non-linearly transformed according to the Divisive Normalization [49]. The DN model has been studied in biological, psychophysical, physiological and perception fields to model the contrast sensitivity function (CSF) [49], masking behavior in perception [50], and to study the neuronal responses in biological visual systems [31,51,52]. In image processing, DN transforms have been exploited in contrast enhancement [53], image compression [54,55], and image quality assessment [56,57]. The literature on human perception has shown a strong relationship between the statistical characteristics of DN transform and the perceptual image quality. The researches on visual perception, like [48,50,58,59], have shown that the DN model can effectively represent the masking phenomena. Moreover, it has been shown that DN model achieves statistical independence and can be represented by Gaussian scale mixture (GSM) model [60]. Gaussian scale mixture (GSM) model has been used in quality assessment techniques using spatial and frequency based DN transforms, e.g., [61–64]. The strong perceptual and statistical relevance of divisive normalization transform advocates its usage in the image quality assessment.

The DNT statistics can effectively capture several image distortion types. In particular, the DNT statistics are sensitive to various kinds of distortion and, more importantly, can be used to estimate the amount of distortion present in the image. In the following, we investigate the impact of various kinds of image distortion on the DNT statistics. To this end, we take a sample image and a set of corrupted versions affected by various distortion models to observe the corresponding impact on the DNT histogram. In Fig. 3, we show a sample image from MCL-3D dataset affected by additive white noise (AWN), Gaussian blur (Gauss), sampling noise (Sample), and JPEG compression distortion (see Fig. 3(b)–(e)). Fig. 3(f) shows the DNT histogram of the original image; Fig. 3(g)–(j) report the DNT histograms of the corrupted images superimposed over the DNT histogram of the original image to show how the histogram varies with the introduction of different types of distortion. One can observe significant changes in the DNT histogram of the impaired images.

Moreover, it can be noted that the shape of the histogram changes differently depending on the distortion type. For example, in presence of AWN distortion, the highest peak in the DNT histogram gets flat and the width of the histogram increases. With Gaussian blur, the DNT histogram values gathers around the center resulting in a decrease in width and an increase in the peak height, which gets even sharper in case of sampling noise. The DNT histogram of the JPEG compressed image comprises a very high peak at the center and few very sharp peaks in its proximity due to quantization. The proposed 3D-IQA algo-

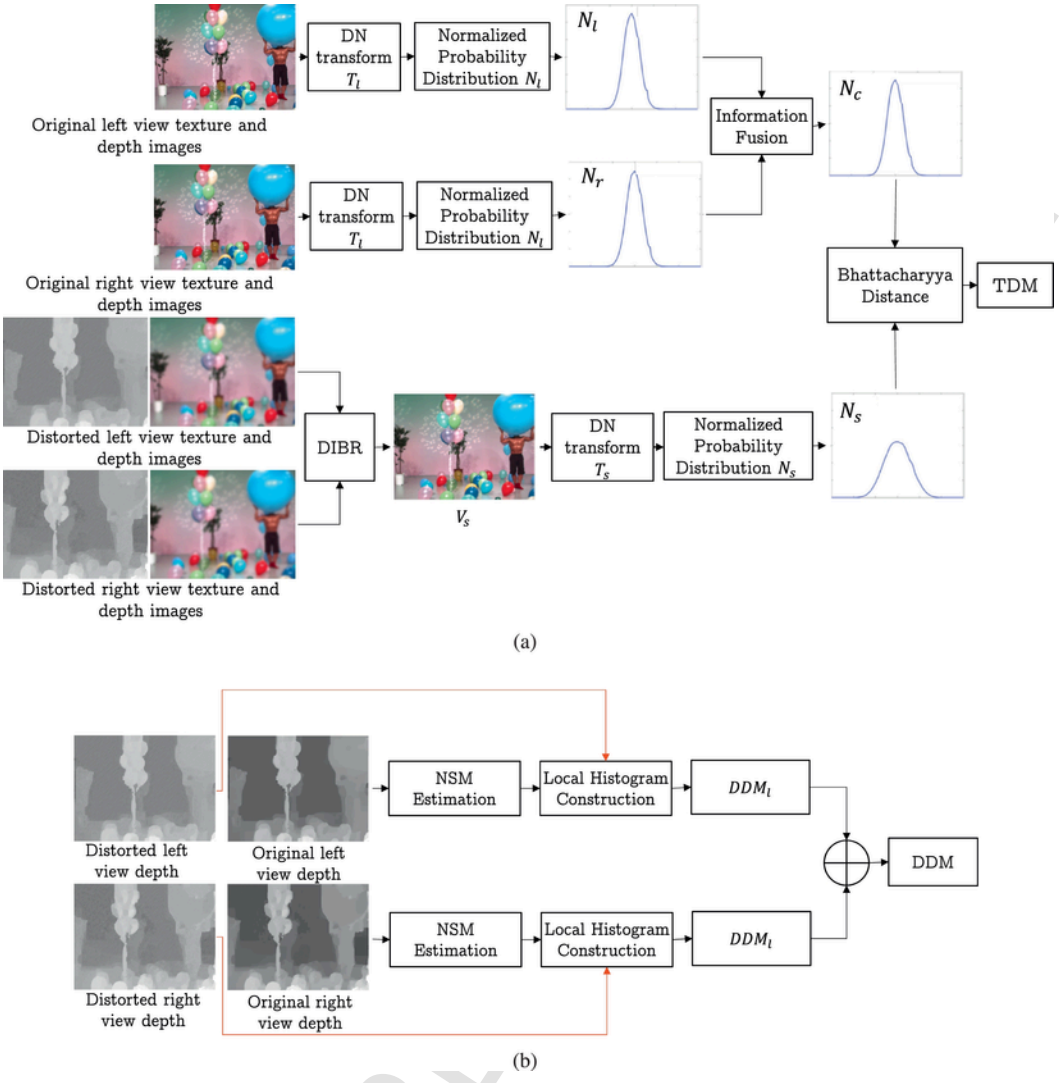

Fig. 2. Overview of the proposed SIQM. Block diagram of TDM (a) and DDM (b).

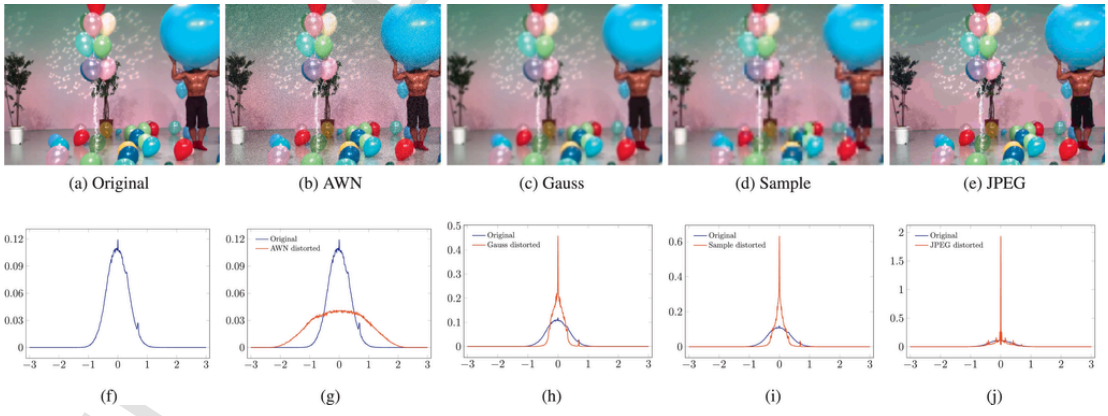

Fig. 3. Histograms of DNT of a sample original image and corrupted versions with different distortion types: (a) sample image of Balloons 3D sequence from MCL-3D dataset, and corresponding DNT histogram (f); (b)–(e) image with AWN, Gaussian, Sampling, and JPEG distortions; (g)–(j) corresponding DTN histograms compared with (f).

rithm aims at exploiting such modifications of the DNT histogram to estimate the quality of DIBR-synthesized images.

3.2. Texture distortion metric (TDM)

Given two views with corresponding depth maps one can compute a virtual intermediate view using DIBR. The cyclopean image is estimated from the reference texture images whereas DIBR uses distorted texture and depth images to obtain the virtual image. As opposed to the theoretical cyclopean image, the virtual image usually suffers from various rendering artifacts (as described in the previous section) which in turn introduce different kinds of distortion limiting its visual quality. The proposed SIQM model aims at leveraging on the discrepancy between these two images to infer the quality of the virtual image.

TDM is based on the estimation of the statistical characteristics of the cyclopean and the virtual image, respectively; those characteristics are compared in order to predict the quality of the virtual image without any reference to the actual intermediate view. In particular, in TDM the color dispersion model based on the divisive normalization (DN) transform is used to characterize images statistically.

TDM takes as input the two reference views V'_l, V'_r and the synthesized view V_s whose quality is subject to assessment. First of all, divisive normalized images T_l, T_r and T_s are created in the spatial domain [64,65]. T_l is computed from V'_l as follows:

$$T_l(u,v) = \frac{V'_l(u,v) - \mu_l(u,v)}{\sigma_l(u,v) + \epsilon} \tag{1}$$

where ϵ is a small constant used to avoid division by zero (in experiments we set $\epsilon = 1$). $\mu_l(u, v)$ and $\sigma_l(u, v)$ are local average and standard deviation computed over a block of size $m \times n$ centered at (u, v). These are computed as:

$$\mu_l(u,v) = \sum_{i=-\frac{m}{2}}^{\frac{m}{2}} \sum_{j=-\frac{n}{2}}^{\frac{n}{2}} w(i,j)\, V'_l(u+i, v+j), \tag{2}$$

$$\sigma_l(u,v) = \sqrt{\sum_{i=-\frac{m}{2}}^{\frac{m}{2}} \sum_{j=-\frac{n}{2}}^{\frac{n}{2}} w(i,j) \left[V'_l(u+i, v+j) - \mu_l(u,v)\right]^2} \tag{3}$$

where w is 2D symmetric Gaussian weight function:

$$w(i,j) = \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}} \tag{4}$$

with $\sigma = 1.16$ as in [65]. The same approach is repeated to get divisive normalized images T_r and T_s using V'_r and V_s, respectively.

The TDM attempts to estimate the statistical characteristics of the cyclopean image from the left and the right DN images T_l and T_r. The existing stereopair quality assessment techniques, e.g., [21,66,67] use the disparity maps and 3D warping to obtain the corresponding cyclopean image; other works, as in [68], perform block based matching over the two stereo images to merge them to obtain a cyclopean image. Unlike previous computationally demanding methods, we propose a simple yet efficient approach to estimate the statistical characteristics of the cyclopean image. We exploit the normalized histograms of DN images of the left and the right texture images to estimate the statistical characteristics of the cyclopean image. Let $\mathcal{N}_l$, $\mathcal{N}_r$ and $\mathcal{N}_s$ be the normalized histograms of T_l, T_r and T_s respectively computed using κ equally spaced bins, i.e. $\sum_{i=1}^{\kappa} \mathcal{N}_{\{l,r,s\}}(i) = 1$. The normalized distribution N_c of the cyclopean image is estimated by fusing the $\mathcal{N}_l$ and $\mathcal{N}_r$ models using the *Levelt weighted sum model* [69,70]:

$$\mathcal{N}_c = w_l \mathcal{N}_l + w_r \mathcal{N}_r \tag{5}$$

where w_l and w_r are weights of $\mathcal{N}_l$ and $\mathcal{N}_r$ respectively such that, $w_l + w_r = 1$. For middle virtual view, we set $w_l = w_r = 0.5$.

In Fig. 4, the left (4(a)) and the right (4(b)) reference stereoscopic images of Balloons sequence from MCL-3D dataset, and the corresponding DN transform images (4(d), (e)) and their normalized distribution models (4(g),4(g)) are shown. It can be observed that the left and right distribution models, whilst very similar, are not identical. In TDM a single distribution $\mathcal{N}_c$ is computed from both the left and the right images and it is considered as representative of the mental image: the resulting distribution is shown in Fig. 5(a). In Fig. 4 we also show a sample intermediate virtual view V_s generated from the stereopair distorted with additive white noise (AWN), along with its corresponding DN transform images (4(f)) and the normalized distribution $\mathcal{N}_s$ (4(i)).

Finally, the distance between the distributions $\mathcal{N}_c$ and $\mathcal{N}_s$ is used as an estimate of the distortion affecting the virtual image. In Fig. 5(b) the two distributions models are superimposed to better notice how they differ in a sample case. The difference between the two curves represents the distortion in the virtual image. To this end we propose to use the *Bhattacharyya coefficient* (ρ) [71]; indeed, this latter has been shown to be more reliable than other metrics, e.g. the Mahalanobis distance [72]. The Bhattacharyya coefficient is used to estimate the similarity between the two distributions as follows:

$$\rho\left(\mathcal{N}_c, \mathcal{N}_s\right) = \sum_{x \in \kappa} \sqrt{\mathcal{N}_c(x)\, \mathcal{N}_s(x)} \tag{6}$$

Finally, TDM is computed as the difference between the two models calculated by the *Hellinger distance* [73]:

$$TDM = \sqrt{1 - \rho\left(\mathcal{N}_c, \mathcal{N}_s\right)} \tag{7}$$

4. Depth distortion metric (DDM)

A depth image is a texture-less grayscale image that usually comprises large homogeneous or linearly changing regions with sharp boundaries. Such depth boundaries, if distorted, cause very annoying structural distortions in the synthesized image. In particular, distortion in depth maps may introduce various artifacts in the synthesized image: blocking artifacts may cause object shifting, ringing around depth edges can result in geometric distortions, etc. [13,14]. Here we propose a *depth distortion metric* (DDM) to estimate the quality of depth images.

The proposed DDM algorithm is based on the fact that the boundary regions in a depth image, being generally very sharp, are more sensitive to noise than the flat homogeneous regions. It exploits statistical tools to assess the distortion present in the boundary regions which determines the quality of the depth map. The DDM works in two steps: first, it computes the *noise sensitivity map* (NSM) of the reference depth image to find the pixels which are the most sensitive to noise, and are expected to introduce significant distortion in the synthesized image. We call such pixels *noise sensitive pixels* (NSP). Second, for each NSP, a local histogram from the distorted depth map is constructed and ana-

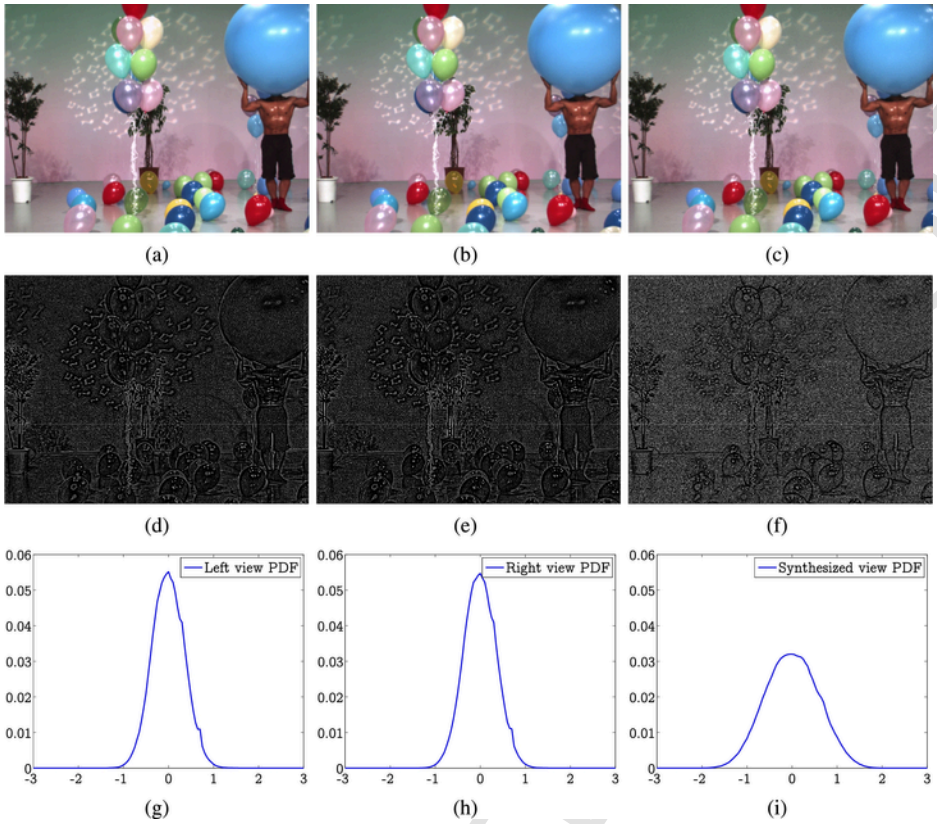

Fig. 4. Sequence: Balloons, Frame: 1, Left View: 1, Right View: 5. Left camera view (a), Right Camera view (b), Synthesized middle view from distorted stereopair (c), the corresponding divisive normalized images (d)–(f) and the normalized probability distributions (g)–(i).

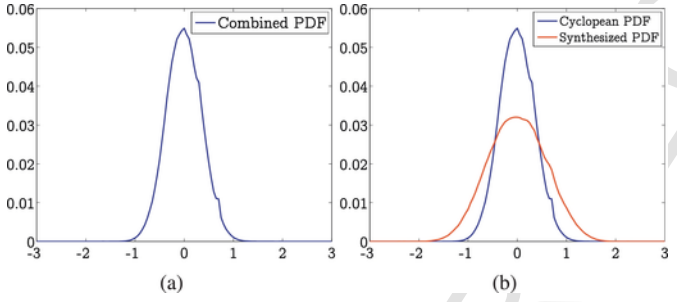

Fig. 5. (a) Probability distribution of estimated cyclopean image. (b) Probability distribution of synthesized view overlaid on (a).

lyzed to estimate the quality of the depth image. The DDM builds on the key observation that distortion in the depth maps significantly alters the histogram around NSPs. The proposed algorithm exploits the shape of the histogram to predict the depth image quality. The following paragraphs describe each step in detail.

To discriminate between the sharp and the homogeneous regions we propose to use the gradient magnitude. Let D'_l be the reference left depth image and D_l be the corresponding distorted depth image (see Fig. 2). The NSM of reference depth map D'_l is computed as

$$NSM = \sqrt{G_x^2 + G_y^2}, \tag{8}$$

where G_x and G_y are the horizontal and vertical gradient components computed with the well-known Sobel filter. The NSM is normalized to bring the values in interval [0, 1]. Moreover, only pixels with $NSM > \tau$ are selected for the DDM estimation.

The distortion of a given NSP is estimated by examining its neighborhood: a local histogram of its neighborhood is constructed and analyzed to infer the presence of noise. As the NSP lies 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 noise, the depth transitions may undergo significant changes, e.g., blurriness, blockiness, and ringing artifacts and this 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. 6 shows histograms of a sample NSP p (with neighborhood of size 15×15) from reference and distorted depth maps of Poznan_Street sequence from MCL-3D dataset (see Section 5 for dataset details). Fig. 6(a) shows the histogram of p from the reference undistorted depth map and the Fig. 6(b)–(l) show the histograms of p when the depth map is distorted with different distortion types and levels. The histogram is computed onto 10 equal bins. Two very high peaks with values above 100 can be observed in Fig. 6(a) showing that the depth values are concentrated around two bins whereas the rest of the histogram is very sparse and almost empty. Fig. 6(b) and (c) shows the histograms of the same region when the depth map is distorted with Gaussian blur at two different levels. The histogram in Fig. 6(b) exhibits lower peaks and a higher valley in between: a drop of more than 30 can be observed in the two peaks along with increased values of the bins in the middle. The increase in Gaussian blur level further smooths the histogram (Fig. 6(c)). A similar trend in histogram can be noted in case of sampling noise (Fig. 6(k), (l)), additive white noise (AWN) (Fig. 6(e), (f)), and under JPEG compression (Fig. 6(g)–(j)); in all cases, when increasing the noise the two high peaks decrease and get distributed over the bins in between. As a consequence, we can exploit the histogram modification to estimate the distortion in the depth image.

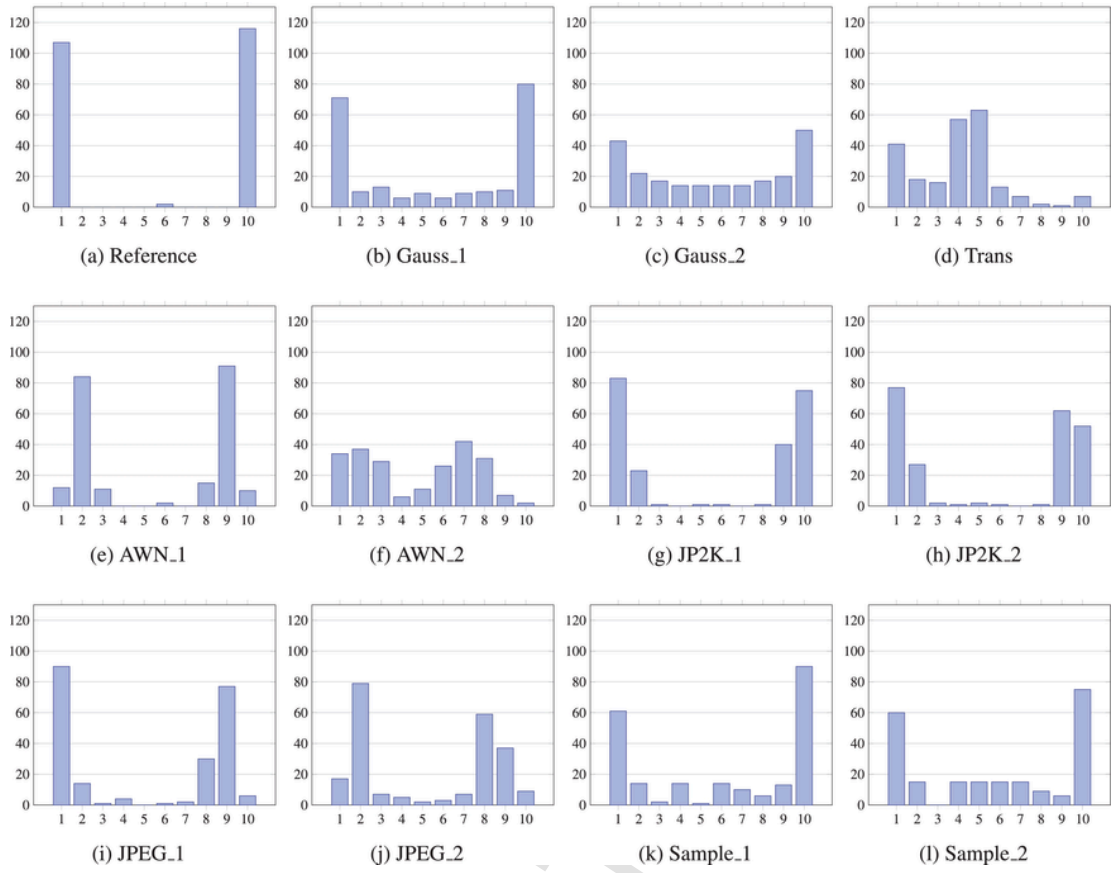

Fig. 6. Histograms of a sample pixel p in the depth image d1 of Poznan_Street sequence from MCL-3D dataset. In (b)–(e) the depth image is distorted with Gaussian blur noise at four different levels. (a) $Q_i = 935$, (b) $Q_i = 575$, (c) $Q_i = 275$, (d) $Q_i = 405$, (e) $Q_i = 605$, (f) $Q_i = 545$, (g) $Q_i = 675$, (h) $Q_i = 565$, (i) $Q_i = 675$, (j) $Q_i = 525$. *Note: in the above graphs, the x-axis represents bin number and the y-axis represents the size of the bin.*

It turns out that the depth distortion can be determined by computing the area lying above the histogram. Such area decreases for increasing noise and can be correlated to the quality of the depth map. Let $\mathcal{S}$ be the set of noise sensitive pixels and let $p_i \in \mathcal{S}$ be an NSP with coordinates $(x, y)|\{1 \le x \le M;\ 1 \le y \le N\}$. For each $p_i \in \mathcal{S}$, we select a patch $\mathcal{P}_i$ in D_l of size $h \times h$ centered at (x, y) and construct its histogram $\mathcal{H}_i^{\kappa}$ computed on κ bins. The quality index $\mathcal{Q}_i$ of pixel p_i is computed by approximating the area contained above the curve $\mathcal{H}_i^{\kappa}$ as follows:

$$\mathcal{Q}_i \approx \sum_{t=1}^{\kappa} \left[\max\left(\mathcal{H}_i^{\kappa}\right) - \mathcal{H}_i^{\kappa}(t)\right] \tag{9}$$

Fig. 6 also shows the Q_i values for each histogram. The Q_i for histogram of the reference depth image (Fig. 6(a)) is very high representing the good quality of the depth map. It can be noted that for each distortion type, Q_i decreases when the noise level is increased. This example clearly shows how Q_i can be related to the amount of distortion that a depth pixel has received. Finally, the DDM of depth image D_l is computed as:

$$DDM_l = \frac{\lambda}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} \frac{1}{\mathcal{Q}_i} \tag{10}$$

where $|\mathcal{S}|$ represents the size of set $\mathcal{S}$ and λ is the scaling factor. In all the following experiments we set $\lambda = 100$. The value of DDM_r for the right depth map D_r is computed analogously. The two distortion measures are averaged with the same weights w_l and w_r used in (5) to estimate the overall depth distortion:

$$DDM = w_l DDM_l + w_r DDM_r \tag{11}$$

4.1. Synthesized image quality metric (SIQM)

The texture distortion (7) and the depth distortion (11) are combined to predict the overall quality of the synthesized picture; the value of the *Synthesized Image Quality Metric* (SIQM) is calculated as:

$$SIQM = [TDM]^{\alpha} \cdot [DDM]^{\beta} \tag{12}$$

where $0 \le \alpha, \beta \le 1$ are the control parameters used to adjust the relative importance of texture and depth distortions. The values of α and β have been empirically set to $\alpha = 0.85$ and $\beta = 0.15$ in this paper.

5. Experimental evaluation and results

In this section, we evaluate the performance of the proposed quality metric on two large DIBR synthesized datasets and compare it with other state-of-the-art synthesized image quality assessment methods.

5.1. Experimental setup and evaluation datasets

The parameter values of TDM and DDM algorithms used in experiments are listed in Table 1; for a fair evaluation in all experiments

Table 1
SIQM parameters settings.

TDM Settings		DDM Settings	
Parameter	Value	Parameter	Value
m, n	7	κ	10
κ	600	τ	0.25

these values were kept unchanged. The values of these parameters are estimated empirically by using few sample texture and depth videos. In case of synthesized stereoscopic images, the SIQM score of each synthesized image is independently computed using (12), and the overall quality of the stereopair is computed by averaging the two SIQM values.

The performance of the proposed SIQM metric is evaluated on MCL-3D dataset [44] and SIAT Synthesized Video Quality dataset (SIAT-3D) [45]. The MCL-3D dataset aims at analyzing the impact of different image distortions on the quality of the DIBR-synthesized images. SIAT-3D on the other hand, focuses on the impact of compression distortion on the virtual images. MCL-3D dataset has been proposed by Media Communications Lab, University of Southern California, and is publicly available.[1] It reports MOS values, of 693 stereopair images generated using DIBR from distorted texture and depth images. The dataset is created from 9 multiview-videos-plus-depth (MVD) sequences. Three views from each test sequence are used and a key frame with corresponding depth maps is selected from each view. Six distortions with four different levels are applied to either texture, depth image or both; the details are presented in Table 2. From these distorted texture images and depth maps, intermediate middle virtual images are generated using *view synthesis reference software* (VSRS) [74] – a depth image based rendering (DIBR) technique. The SIAT-3D[2][45] dataset consists of synthesized videos from 10 different MVD sequences (1024×768 and 1920×1088 resolution). Two viewpoints of each sequence are compressed with 3DV-ATM v10.0 with 14 different texture/depth quantization combinations. From these compressed texture and depth videos virtual views were generated by using the VSRS-1D-Fast software implemented in the recent 3D-HEVC reference software [11]. Thus a total of 140 synthesized videos are available. These videos were rated by 40 subjects using single stimulus paradigm with continuous score to obtain subjective MOS values.

5.2. Performance evaluation parameters

For performance evaluation according to the VQEG guidelines [75], we use Pearson linear correlation coefficient (PLCC) as *prediction accuracy* test and the Spearman rank order correlation coefficient (SROCC) as *prediction monotonicity* test. Root Mean Square Error (RMSE) measure is used to estimate the *prediction error* in objective scores and the subjective ratings. Before computing these parameters, as recommended by the video quality expert group (VQEG) [75], the predicted scores are mapped to the subjective ratings with a monotonic nonlinear regression. The logistic function outlined in [76] is used for regression mapping:

[1] http://www.mcl.usc.edu/mcl-3d-database/.
[2] http://codec.siat.ac.cn/SIATDatabase/index.html.

Table 2
Distortion types and control parameters with range values used in experimental evaluation.

Distortion	Control parameter	Parameter values
Gaussian Blur (Gauss)	Standard Deviation	11, 21, 31, 41
Additive White Noise (AWN)	Standard Deviation	5, 17, 33, 53
Down-sampling Blur (Sample)	Sampling Ratio	5, 8, 11, 14
JPEG Compression (JPEG)	Quality Level	30, 12 , 8, 5
JPEG2000 Compression (JP2K)	Compression Param.	200, 500, 900, 1500
Transmission Loss (Trans)	OpenJPEG library	Visually selected

$$Q_p = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp\left(\beta_2 \left(Q - \beta_3\right)\right)} \right) + \beta_4 Q + \beta_5 \tag{13}$$

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

As described earlier, in evaluating the quality of synthesized images the proposed 3D-IQA metric does not require the corresponding reference images hence, it is not a full-reference IQA metric. For performance evaluation, we compared SIQM with the widely used 3D-IQA metrics: 3DSwIM [34], You$_l$ (local model) [32], You$_g$ (global model) [32], StSD [25], Benoit [37], Gorley [29], ST-SIAQ [35], and NIQSV [36]. In this set of compared methods, NIQSV is a no-reference technique while the rest are full-reference quality metrics.

5.3. Performance evaluation on MCL-3D and SIAT-3D datasets

The overall evaluation results of SIQM and other 3D -IQA algorithms on MCL-3D dataset in terms of PLCC, SROCC and RMSE are presented Table 3. The statistics presented in the table show that in all three performance criteria SIQM outperforms all the compared methods. The performance of the proposed SIQM algorithm and other 3D-IQA algorithms is also evaluated on individual distortions listed in Table 2. Tables 4–6 show the results of SIQM and other 3D-IQA algorithms on individual distortion types in terms of PLCC, SROCC and RMSE, respectively. We conclude from these results that on individual distortions SIQM generally outperforms most other 3D-IQA algorithms. In particular, in terms of PLCC SIQM performs the best on Gauss, Sample, and JPEG distortions. You and Gorely algorithms perform better than SIQM on Transloss and JP2K distortions respectively. Similar statistics can be observed in SROCC and RMSE performance parameters reported in Tables 5 and 6).

We also present the performance analysis of the proposed 3D quality metric on SIAT Synthesized Video Quality dataset (SIAT-3D) [45]. SIAT-3D has been designed to test the performance of 3D-IQA algorithms in presence of 3D video compression distortion. To measure the performance of SIQM we used 50 test sequences where both texture and depth are compressed with different quantization parameters and we omit the limit cases where either texture or depth videos are left uncompressed. Table 7 reports the performance of SIQM and other 3D-IQA algorithms.

From these results, we can note that the proposed algorithm outperforms all compared methods with average PLCC of 0.658, SROCC of 0.6185, and 0.0846 RMSE.

Table 3
Overall performance of SIQM and other 3D-IQA metrics on MCL-3D dataset.

Metric	3DSwIM	You_l	You_g	StSd	Benoit	Gorley	ST-SIAQ	NIQSV	SIQM
PLCC	0.6497	0.7504	0.3650	0.6995	0.7425	0.7099	0.7133	0.6783	**0.7744**
SROCC	0.5683	0.7567	0.6609	0.7008	0.7518	0.7196	0.7034	0.6208	**0.7756**
RMSE	1.9777	1.7196	2.4222	1.8593	1.7429	1.8323	1.8233	1.9118	**1.6461**

Table 4
Comparison of SIQM and other 3D-IQA algorithms on different distortion types in terms of PLCC.

Metric	AWN	Gauss	Sample	Transloss	JPEG	JP2K
3DSwIM	0.4640	0.8218	0.8127	0.7566	0.6431	0.6478
StSD	0.7472	0.8429	0.8392	0.6527	0.7372	0.7836
Benoit	0.9102	0.8600	0.8544	0.6796	0.8044	0.8064
You_l	**0.9278**	0.8560	0.8564	**0.7930**	0.7934	0.8071
You_g	0.8856	0.8685	0.7138	0.5373	0.8599	0.8921
Gorley	0.7735	0.8550	0.8544	0.6043	0.8326	**0.9051**
ST-SIAQ	0.7884	0.8215	0.8070	0.8264	0.7420	0.7884
NIQSV	0.8184	0.8356	0.8262	0.7190	0.2798	0.7070
SIQM	0.7798	**0.8738**	**0.8673**	0.6851	**0.8858**	0.7057

Table 5
Comparison of SIQM and other 3D-IQA algorithms on different distortion types in terms of SROCC.

Metric	AWN	Gauss	Sample	Transloss	JPEG	JP2K
3DSwIM	0.4560	0.6999	0.7247	0.8134	0.7029	0.5526
StSD	0.7242	0.7993	0.8146	0.6290	0.7663	0.7896
Benoit	0.9116	0.8222	0.8354	0.6809	0.8200	0.8143
You_l	0.9214	0.8115	0.8256	0.8084	0.8115	0.8098
You_g	0.8948	0.8555	0.7825	0.5785	0.8085	0.8561
Gorley	0.7841	0.7915	0.8218	0.5800	**0.8345**	**0.8910**
ST-SIAQ	0.7765	0.7235	0.7668	**0.8408**	0.6638	0.8092
NIQSV	0.8272	0.8447	0.8478	0.7461	0.2753	0.5701
SIQM	**0.9278**	**0.8560**	**0.8564**	0.7930	0.7934	0.8071

Table 6
Comparison of SIQM and other 3D-IQA algorithms on different distortion types in terms of RMSE.

Metric	AWN	Gauss	Sample	Transloss	JPEG	JP2K
3DSwIM	2.1966	1.5263	1.6953	1.3098	1.7564	1.8770
StSD	1.6479	1.4412	1.5820	1.5177	1.5496	1.5308
Benoit	1.6268	1.3666	1.5117	1.4695	1.3627	1.4570
You_l	1.5621	1.3848	1.5019	1.2203	1.3960	1.4549
You_g	1.6518	1.3276	2.0377	1.6895	1.1707	1.6099
Gorley	1.5718	1.3891	1.5118	1.5961	1.2702	**1.0479**
ST-SIAQ	1.5255	1.5275	1.7180	**1.1281**	1.5377	1.5158
NIQSV	1.6250	1.4771	1.5479	1.6926	2.2020	1.7726
SIQM	**1.5524**	**1.3027**	**1.4481**	1.4592	**1.0642**	1.7457

Table 7
Overall performance of SIQM and other 3D-IQA metrics on SIAT-synthesized video quality dataset.

Metric	3DSwIM	You_l	You_g	StSD	Benoit	Gorley	ST-SIAQ	NIQSV	SIQM
PLCC	0.2275	0.5230	0.4427	0.6517	0.6423	0.6464	0.3989	0.6248	**0.6585**
SROCC	0.2394	0.4634	0.4332	0.6051	0.6179	0.6109	0.3626	0.5445	**0.6185**
RMSE	0.1094	0.0958	0.1007	0.0852	0.0861	0.0857	0.1030	0.0877	**0.0846**

5.4. Statistical significance test

In order to draw statistically meaningful conclusions about the performance of the proposed quality metric, we conducted statistical significant tests. We test the Gaussianity of the residual differences between the DMOS and 3D-IQA predictions after non-linear mapping, and use the F-statistic to compare the variance of the two distributions [76]. In particular the F-test is used to check if, under Gaussian distribution hypothesis, the residuals of two quality metrics being evaluated come from the same distribution and therefore are statistically indistinguishable [45,76,77]. The ratio between variances of the resid-

uals of the two quality metrics is computed and compared with the F-ratio to determine their significance. The F-ratio threshold is obtained from the F-distribution look-up table with *alpha* = 0.05.

The results of statistical significance test are presented in Table 8. Each entry in the table is a codeword of 8 characters. The first six symbols correspond to the distortion types AWN, Gauss, Sample, Transloss, JPEG, and JP2K of MCL-3D dataset, respectively (Table 2). The seventh symbol correspond to MCL-3D whole dataset, and the eighth symbol represents the SIAT-3D dataset. Symbol '1' in the codeword means that the IQA algorithm in row is statistically better than that on the column and symbol '0' that the quality metric on the column is better than that on the row. The symbol '–' denotes that the two quality assessment algorithms are statistically indistinguishable. It can be noted from the reported analysis that the proposed SIQM performs the best among all compared methods on subsets AWN, Gauss, Sample, and JPEG. However, in Transloss and JP2K, You_l and Gorley are statistically better metrics, respectively. We see that on the whole MCL-3D dataset, SIQM performs significantly better than other algorithms, except Benoit and Gorley which performs equivalently well. On SIAT-3D dataset, our method performs significantly better than all the competing 3D-IQA algorithms.

The results of performance evaluation on MCL-3D and SIAT-3D datasets allow us to conclude that the performance achieved by the proposed SIQM is very convincing, considering that it evaluates the quality of DIBR-synthesized image in absence of the corresponding reference images.

6. Conclusion

In this paper, a 3D-IQA algorithm has been presented to estimate the quality of DIBR generated virtual images. We proposed two metrics, one to estimate the quality of the synthesized image and the other to compute the quality of the depth map; the two metrics are then combined into a single quality indicator: SIQM. The texture distortion metric is based upon the cyclopean eye theory and divisive normalization. It estimates the statistical characteristics of the cyclopean images from the input original and the distorted images. Then these characteristics are compared to assess the quality of the virtual image without using the original image corresponding to the virtual image. In the second metric, we use the local histograms of the depth maps to evaluate their quality. The experimental evaluation of the proposed 3D-IQA algorithm carried out on two 3D synthesized image datasets has shown its effectiveness.

Table 8
Statistical significance matrix between DMOS and the metric predicted quality scores on MCL-3D dataset. A value of '1' means that the performance of the quality metric in the row is statistically better than that of the column. A value of '0' indicates that the performance of the quality metric in the row is statistically worse than that of the column and '–' means that the two metrics are statistically equivalent. The first 6 symbols in the codeword represent the distortion types in MCL-3D dataset in the following order: AWN, Gauss, Sample, Transloss, JPEG, and JP2K. The seventh and eighth symbol corresponds to the MCL and SIAT datasets, respectively.

Metric	3DSwIM	You_l	You_g	StSd	Benoit	Gorely	ST-SIAQ	NIQSV	SIQM
3DSwIM	——	0000000-	0111001-	0-0-00-0	000-0000	00010000	0—000-	000-10-	000-0000
You_l	1111111-	——	-1110-1-	1-11-1-	0-1—-	—100-	-1111—	—1111-	—10-0
You_g	1000110-	-0001-0-	——	1000110-	-0001-0-	-00-100-	-000110-	-000110-	-0000-00
StSd	1-1-11-1	0-00-0-	0111001-	——	0—0-0-	—-00-	0-1—-	—-11-	-00-000-
Benoit	111-1111	1-0—-	-1110-1-	1—1-1-	——	1-100-	-11-1—	1—111-	1—0-0
Gorely	11101111	—011-	-11-011-	—-11-	0-011-	——	-11-11-	—011-	—00100
ST-SIAQ	1—111-	-0000—	-111001-	1-0—-	-00-0—	-00-00-	——	-00-11-	-00-0-00
NIQSV	111-01-	—0000-	-111001-	—-00-	0—000-	—100-	-11-00-	——	—-0000
SIQM	111-1111	—01-1	-1111-11	-11-111-	0—1-1	—11011	-11-1-11	—-1111	——

References

[1] M. Tanimoto, FTV: free-viewpoint television, Signal Process.-Image Commun. 27 (6) (2012) 555–570.

[2] M. P. Tehrani, T. Senoh, M. Okui, K. Yamamoto, N. Inoue, T. Fujii, H. Nakamura, Proposal to consider a new work item and its use case - rei : an ultra-multiview 3D display. ISO/IEC JTC1/SC29/WG11/m30022.

[3] A. Boev, D. Hollosi, A. Gotchev, K. Egiazarian, Classification and simulation of stereoscopic artifacts in mobile 3dtv content, Proceedings of SPIE, 7237, 2009. 72371F–72371F–12.

[4] L. Xing, J. You, T. Ebrahimi, A. Perkis, Assessment of stereoscopic crosstalk perception, IEEE Trans. Multimedia 14 (2) (2012) 326–337.

[5] Q. Huynh-Thu, P.L. Callet, M. Barkowsky, Video quality assessment: from 2d to 3d - challenges and future trends, Proceedings 17th IEEE International Conference Image Processing (ICIP), 20104025–4028.

[6] A. Boev, M. Poikela, A. Gotchev, A. Aksay, Modelling of the stereoscopic HVS. Report on Mobile 3DTV {http://www.sp.cs.tut.fi/mobile3dtv/results/}.

[7] E. Bosc, M. Koppel, R. Pepion, M. Pressigout, L. Morin, P. Ndjiki-Nya, P.L. Callet, Can 3d synthesized views be reliably assessed through usual subjective and objective evaluation protocols?, Proc. 18th IEEE Int. Conf. Image Process. (ICIP) (2011) 2597–2600.

[8] E. Bosc, P. Hanhart, P.L. Callet, T. Ebrahimi, A quality assessment protocol for free-viewpoint video sequences synthesized from decompressed depth data, Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on, 2013100–105.

[9] A.K. Moorthy, C.-C. Su, A. Mittal, A.C. Bovik, Subjective evaluation of stereoscopic image quality, Signal Process.-Image Commun. 28 (8) (2013) 870–883.

[10] C. Fehn, Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV, Proc. SPIE 5291 (2004) 93–104, https://doi.org/10.1117/12.524762.

[11] K. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. Rhee, G. Tech, M. Winken, T. Wiegand, 3D high-efficiency video coding for multi-view video and depth data, IEEE Trans. Image Process. 22 (9) (2013) 3366–3378.

[12] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vis. 47 (1–3) (2002) 7–42, https://doi.org/10.1023/A:1014573219977.

[13] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Mueller, P. de With, T. Wiegand, The effects of multiview depth video compression on multiview rendering, Signal Process.-Image Commun. 24 (1) (2009) 73–88.

[14] E. Bosc, P.L. Callet, L. Morin, M. Pressigout, Visual quality assessment of synthesized views in the context of 3D-TV, in: C. Zhu, Y. Zhao, L. Yu, M. Tanimoto (Eds.), 3D-TV System with Depth-Image-Based Rendering, Springer, New York, 2013, pp. 439–473, https://doi.org/10.1007/978-1-4419-9964-1_15.

[15] M.S. Farid, M. Lucenteforte, M. Grangetto, Edge enhancement of depth based rendered images, Proceedings of 21st IEEE International Conference Image Processing (ICIP), 20145452–5456.

[16] C. Hewage, M. Martini, Quality of experience for 3D video streaming, IEEE Commun. Mag. 51 (5) (2013) 101–107, https://doi.org/10.1109/MCOM.2013.6515053.

[17] M.-J. Chen, A.C. Bovik, L.K. Cormack, Study on distortion conspicuity in stereoscopically viewed 3d images, IVMSP Workshop, IEEE, 201124–29.

[18] P.-C. Wang, S.-L. Hwang, H.-Y. Huang, C.-F. Chuang, System cross-talk and three-dimensional cue issues in autostereoscopic displays, J. Electron. Imaging 22 (1) (2013). 013032–013032.

[19] W.J. Tam, L.B. Stelmach, P.J. Corriveau, Psychovisual aspects of viewing stereoscopic video sequences, Proceedings of SPIE, 3295, 1998226–235.

[20] W. Chen, J. Fournier, M. Barkowsky, P.L. Callet, Quality of experience model for 3dtv, Proceedings of SPIE, 8288, 20121–9.

[21] M.-J. Chen, C.-C. Su, D.-K. Kwon, L.K. Cormack, A.C. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Process.-Image Commun. 28 (9) (2013) 1143–1155.

[22] Y.-H. Lin, J.-L. Wu, Quality assessment of stereoscopic 3d image compression by binocular integration behaviors, IEEE Trans. Image Process. 23 (4) (2014) 1527–1542, https://doi.org/10.1109/TIP.2014.2302686.

[23] F. Shao, K. Li, W. Lin, G. Jiang, Q. Dai, Learning blind quality evaluator for stereoscopic images using joint sparse representation, IEEE Trans. Multimedia (2016) 1. 99

[24] F. Qi, D. Zhao, W. Gao, Reduced reference stereoscopic image quality assessment based on binocular perceptual information, IEEE Trans. Multimedia 17 (12) (2015) 2338–2344.

[25] V. De Silva, H. Arachchi, E. Ekmekcioglu, A. Kondoz, Toward an impairment metric for stereoscopic video: a full-reference video quality metric to assess compressed stereoscopic video, IEEE Trans. Image Process. 22 (9) (2013) 3392–3404, https://doi.org/10.1109/TIP.2013.2268422.

[26] J. Yang, Y. Wang, B. Li, W. Lu, Q. Meng, Z. Lv, D. Zhao, Z. Gao, Quality assessment metric of stereo images considering cyclopean integration and visual saliency, Inf. Sci. (Ny) 373 (2016) 251–268.

[27] E. Bosc, P.L. Callet, L. Morin, M. Pressigout, An edge-based structural distortion indicator for the quality assessment of 3D synthesized views, Proceedings of Picture Coding Symposium (PCS), 2012249–252.

[28] H. Shao, X. Cao, G. Er, Objective quality assessment of depth image based rendering in 3dtv system, Proceedings of 3DTV Conference, True Vision,-Capture Transmission Display 3D Video (3DTV-CON), 20091–4.

[29] P. Gorley, N. Holliman, Stereoscopic image quality metrics and compression, Proceedings of SPIE, 6803, 2008. 680305–680305–12.

[30] D. Kim, S. Ryu, K. Sohn, Depth perception and motion cue based 3D video quality assessment, Proceedings of IEEE International Symposium on Broadband Multimedia Systems Broadcast. (BMSB), 20121–4.

[31] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612, https://doi.org/10.1109/TIP.2003.819861.

[32] J. You, L. Xing, A. Perkis, X. Wang, Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis, Proceedings of International Workshop Video Processing Quality Metrics Consumer Electronics, 20101–5.

[33] C.-T. Tsai, H.-M. Hang, Quality assessment of 3D synthesized views with depth map distortion, Proceedings of International Conference on Visual Communications and Image Processing (VCIP), 20131–6, https://doi.org/10.1109/VCIP.2013.6706348.

[34] F. Battisti, E. Bosc, M. Carli, P.L. Callet, S. Perugia, Objective image quality assessment of 3d synthesized views, Signal Process. Image Commun. 30 (0) (2015) 78–88.

[35] S. Ling, P.L. Callet, Image quality assessment for free viewpoint video based on mid-level contours feature, Proceedings of IEEE International Conference on Multimedia and Expo (ICME), 201779–84.

[36] S. Tian, L. Zhang, L. Morin, O. Deforges, NIQSV: a no reference image quality assessment metric for 3D synthesized views, Proceedings of Int Conf. Acoustics, Speech and Signal Process. (ICASSP), 20171248–1252.

[37] A. Benoit, P.L. Callet, P. Campisi, R. Cousseau, Quality assessment of stereoscopic images, EURASIP J. Image Video Process. 2008 (1) (2009). 659024.

[38] M. Carnec, P.L. Callet, D. Barba, An image quality assessment method based on perception of structural information, Proceedings of International Conference Image Processing (ICIP), 3, 2003. III–185–8 vol.2.

[39] P.-H. Conze, P. Robert, L. Morin, Objective view synthesis quality assessment, Proceedings of SPIE, 8288, 2012. 82881M–82881M–14.

[40] S. Winkler, D. Min, Stereo/multiview picture quality: overview and recent advances, Signal Process.-Image Commun. 28 (10) (2013) 1358–1373, https://doi.org/10.1016/j.image.2013.07.008.

[41] A.K. Moorthy, A.C. Bovik, A survey on 3d quality of experience and 3d quality assessment, Proceedings of SPIE, 8651, 2013https://doi.org/10.1117/12.2008355. 86510M–86510M–11.

[42] M.S. Farid, M. Lucenteforte, M. Grangetto, Objective quality metric for 3D virtual views, Proceedings of International Conference on Image Process. (ICIP), 20153720–3724.

[43] M.S. Farid, M. Lucenteforte, M. Grangetto, Blind depth quality assessment using histogram shape analysis, Proceedings of 3DTV Conference on True Vision,-Capture Transmission Display 3D Video (3DTV-CON), 20151–5.

[44] R. Song, H. Ko, C. Kuo, MCL-3d: a database for stereoscopic image quality assessment using 2d-image-plus-depth source, Journal of Information Science and Engineering, 31, 20151593–1611.

[45] X. Liu, Y. Zhang, S. Hu, S. Kwong, C.-C. Kuo, Q. Peng, Subjective and objective video quality assessment of 3D synthesized views with texture/depth compression distortion, IEEE Trans. Image Process. 24 (12) (2015) 4847–4861, https://doi.org/10.1109/TIP.2015.2469140.

[46] B. Julesz, Cyclopean perception and neurophysiology, Invest. Ophthalmol. Vis. Sci. 11 (6) (1972) 540–548.

[47] D.J. Heeger, Normalization of cell responses in cat striate cortex, Visual Neurosci. 9 (02) (1992) 181–197.

[48] P. Teo, D. Heeger, Perceptual image distortion, Proceedings of International Conference on Image Process. (ICIP), 2, 1994982–986.

[49] V. Laparra, J.M.n. Marí, J. Malo, Divisive normalization image quality metric revisited, J. Opt. Soc. Am. A 27 (4) (2010) 852–864.

[50] J.M. Foley, Human luminance pattern-vision mechanisms: masking experiments require a new model, J. Opt. Soc. Am. A 11 (6) (1994) 1710–1719.

[51] I. Epifanio, J. Gutirrez, J. Malo, Linear transform for simultaneous diagonalization of covariance and perceptual metric matrix in image coding, Pattern Recognit. 36 (8) (2003) 1799–1811.

[52] E.P. Simoncelli, D.J. Heeger, A model of neuronal responses in visual area mt, Vision Res. 38 (5) (1998) 743–761.

[53] Z. Wang, E.P. Simoncelli, Translation insensitive image similarity in complex wavelet domain, Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2, 2005573–576.

[54] S. Lyu, Divisive normalization: Justification and effectiveness as efficient coding transform, Advances in Neural Information Processing Systems, 20101522–1530.

[55] J. Malo, I. Epifanio, R. Navarro, E.P. Simoncelli, Nonlinear image representation for efficient perceptual coding, IEEE Trans. Image Process 15 (1) (2006) 68–80.

[56] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE J. Sel. Topics Signal Process. 3 (2) (2009) 202–211.

[57] X. Wang, G. Jiang, M. Yu, Reduced reference image quality assessment based on contourlet domain and natural image statistics, 5th International Conference on Image and Graphics, 200945–50.

[58] A.B. Watson, J.A. Solomon, Model of visual contrast gain control and pattern masking, J. Opt. Soc. Am. A 14 (9) (1997) 2379–2391.
[59] E.P.S.S. Lyu, Statistically and perceptually motivated nonlinear image representation, Proceedings of SPIE, 6492, 2007. 6492–6492–15.
[60] S. Lyu, E. Simoncelli, Nonlinear image representation using divisive normalization, Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 20081–8, https://doi.org/10.1109/CVPR.2008.4587821.
[61] X. Wang, G. Jiang, M. Yu, Reduced reference image quality assessment based on contourlet domain and natural image statistics, Proceedings of 5th International Conference Image Graph. (ICIG), 200945–50, https://doi.org/10.1109/ICIG.2009.44.
[62] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE J. Sel. Topics Signal Process. 3 (2) (2009) 202–211, https://doi.org/10.1109/JSTSP.2009.2014497.
[63] A. Mittal, R. Soundararajan, A. Bovik, Making a completely blind image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212, https://doi.org/10.1109/LSP.2012.2227726.
[64] A. Mittal, A. Moorthy, A. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[65] D. Ruderman, The statistics of natural images, Netw.-Comput. Neural Syst. 5 (4) (1994) 517–548.
[66] M.-J. Chen, L. Cormack, A. Bovik, No-reference quality assessment of natural stereopairs, IEEE Trans. Image Process. 22 (9) (2013) 3379–3391.
[67] R. Akhter, Z.M.P. Sazzad, Y. Horita, J. Baltes, No-reference stereoscopic image quality assessment, Proceedings of SPIE, 7524, 2010. 75240T–75240T–12.10.1117/12.838775.
[68] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, G. Akar, Towards compound stereo-video quality metric: a specific encoder-based framework, Proceedings of IEEE Southwest Symp. Image Analysis Interpretation, 2006218–222.
[69] W. Levelt. On binocular rivalry, Soesterberg (The Netherlands): Institute for Perception RVO-TNO,| c1965 1.
[70] S. Grossberg, F. Kelly, Neural dynamics of binocular brightness perception, Vision Res. 39 (22) (1999) 3796–3816.
[71] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distributions, Bull. Calcutta Math. Soc. 35 (1943) 99–109.
[72] P.C. Mahalanobis, On the generalised distance in statistics, Proceedings National Institute of Science, India, 2, 193649–55.
[73] E. Hellinger, Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen., J. Reine Angew. Math. 136 (1909) 210–271.
[74] M. Tanimoto, T. Fujii, K. Suzuki. View synthesis reference software (VSRS) 3.5 (Mar 2010).ftp://ftp.merl.com/pub/avetro/3dv-cfp/.
[75] 2003. Video Quality Expert Group, Final report from the video quality experts group on the validation of objective models of video quality assessment.
[76] H. Sheikh, M. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3440–3451, https://doi.org/10.1109/TIP.2006.881959.
[77] D.E. Moreno-Villamarin, H.D. Benitez-Restrepo, A.C. Bovik, Predicting the quality of fused long wave infrared and visible light images, IEEE Trans. Image Process. 26 (7) (2017) 3479–3491.