Perceptual quality assessment of 3D synthesized images

This is the author's manuscript

Original Citation:

Availability:
This version is available http://hdl.handle.net/2318/1651383 since 2018-01-22T11:05:25Z

Published version:
DOI:10.1109/ICME.2017.8019307

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.
PERCEPTUAL QUALITY ASSESSMENT OF 3D SYNTHESIZED IMAGES

Muhammad Shahid Farid∗†, Maurizio Lucenteforte†, Marco Grangetto†

†Dipartimento di Informatica, Università degli Studi di Torino, Torino, Italy
*College of Information Technology, University of the Punjab, Lahore, Pakistan

ABSTRACT
Multiview video plus depth (MVD) is the most popular 3D video format where the texture images contain the color information and the depth maps represent the geometry of the scene. The depth maps are exploited to obtain intermediate views to enable 3D-TV and free-viewpoint applications using the depth image based rendering (DIBR) techniques. DIBR is used to get an estimate of the intermediate views but has to cope with depth errors, occlusions, imprecise camera parameters, re-interpolation, to mention a few issues. Therefore, being able to evaluate the true perceptual quality of synthesized images is of paramount importance for a high quality 3D experience. In this paper, we present a novel algorithm to assess the quality of the synthesized images in the absence of the corresponding references. The algorithm uses the original views from which the virtual image is generated to estimate the distortion induced by the DIBR process. In particular, a block-based perceptual feature matching based on signal phase congruency metric is devised to estimate the synthesis distortion. The experiments worked out on standard DIBR synthesized database show that the proposed algorithm achieves high correlation with the subjective ratings and outperforms the existing 3D quality assessment algorithms.

Index Terms— Quality assessment, depth image based rendering, view synthesis, Free-viewpoint TV

1. INTRODUCTION
Two slightly different views are projected simultaneously in 3D television (3DTV), each filtered to the respective eye to achieve depth sensation. The recent 3D video applications e.g., free-viewpoint television (FTV) [1] can provide smooth horizontal parallax by presenting a different set of views to the viewer depending on his position. Multiview video plus depth (MVD) format [2] has been adopted by the Moving Picture Experts Group (MPEG) for 3DTV and future FTV technologies [3]. MVD, in addition to texture images, provides per pixel depth values of the color image which allows the generation of novel views by using depth image based rendering (DIBR) technique [4–8]. The depth maps are noisy and imperfect as they are usually estimated using stereo matching algorithms. When used in DIBR, they may introduce various textural and structural distortions in the virtual image [9–11].

Quality assessment of synthesized images is an important task in 3DTV framework. Rendered images suffer from different types of artifacts, and the conventional 2D image quality assessment (2D-IQA) algorithms are found to be inappropriate [12, 13]. The problem becomes even more challenging in absence of the reference images corresponding to the virtually generated views. Indeed in many practical cases, it is unpractical to shoot, compress and deliver the set of different views required by a modern immersive display: in such a case many of the images that will be viewed do not have a correspondent reference (ground truth).

Unfortunately, most existing quality assessment algorithms for synthesized image are full reference. Therefore, the quality DIBR-synthesized views must be evaluated in the absence of the corresponding reference views is momentous for a satisfactory quality of experience (QoE). In this paper, we propose a novel algorithm to estimate the quality of the DIBR synthesized images with the following characteristics:

• The quality of the synthesized image is estimated using only the left and (or) the right views from which the virtual view is generated. Thus the proposed algorithm can work when the original reference of the picture under evaluation is not available;

• Signal phase congruency [14] is exploited to devise a quality metric that targets 3D distortion artifacts affecting in particular edges, lines and corners that are considered as important perceptual features;

• The proposed metric yields very high correlation with subjective scores when evaluated on standard DIBR image database.

The rest of the paper is organized as follows: in Sect. 2 the related work is briefly presented; the proposed quality metric is described in Sect. 3, followed by experimental evaluation in Sect. 4. The conclusions are drawn in Sect. 5.

2. RELATED WORK
Due to the increasing popularity of 3DTV, in the recent years a number of algorithms to objectively assess the quality of
the synthesized and stereoscopic images have been proposed. Most of these metrics extend the conventional 2D-IQA algorithms to 3D data. Bosc et al. [15] proposed to compare the edges of the synthesized and the original images to estimate the quality of the DIBR synthesized image. This metric however, is limited to structural distortion and does not consider the color distortion in evaluating the image quality. Perceptual quality metric (PQM) [16] estimates the quality of the synthesized video from the luminance and contrast distortions. The CSED algorithm [17] compares the synthesized and the reference images to estimate the color distortion from the hole regions and structural distortion from the edge sharpness of the reference and virtual images. The algorithm in [18] uses the Hausdorff distance to estimate the structural distortion in the synthesized image and combines it with SSIM score to obtain the final quality score. 3VQM metric [19] estimates the ideal depth map and compares it with the given depth map to estimate the temporal and spatial variations, which are used to estimate the quality of the synthesized video. The 3DSwIM algorithm [20] compares the statistical features of wavelet subbands of the synthesized and the reference images to estimate the quality. Moreover, based on the assumption that viewers are more sensitive to distortion around human silhouettes, in [20] a skin detector is used as an approximate attention model.

The 3D-IQA algorithm proposed in [21] estimates the contrast and luminance distortions by comparing the high spatial frequency regions of the reference and distorted stereopair images. Synthesized image quality evaluator (SIQE) [22], a reduced-reference 3D-IQA, exploits the cyclopean eye theory and divisive normalization to estimate statistical characteristics of the synthesized image and the cyclopean image which are compared to predict the quality of virtual image. View synthesis quality assessment (VSQA) algorithm [23] combines SSIM with a weighting map which is computed from the contrast, orientation and texture maps of the reference and the synthesized images. A blind quality assessment algorithm to assess the compression distortion in the depth maps is proposed in [24]. A morphological wavelet-based peak signal-to-noise ratio measure (MW-PSNR) is proposed in [25] which estimates the structural distortion in the synthesized image to assess its quality. A similar multi-scale quality metric based on morphological pyramids is proposed in [26].

3. PROPOSED PERCEPTUAL QUALITY ASSESSMENT METRIC FOR SYNTHESIZED IMAGES

Depth image based rendering (DIBR) is used to generate images at novel viewpoints in multiview video plus depth (MVD) format. DIBR obtains novel views by warping the nearest available pair of images to the virtual viewpoint with the help of the corresponding depth maps. As described in the previous section, the depth maps being noisy and inaccurate introduce various types of structural and textural distortions in the synthesized image. Evaluating the quality of the synthesized images is an important activity in 3DTV framework, and it becomes a challenging task when the corresponding reference views are unavailable. The proposed quality assessment algorithm estimates the quality of the synthesized image from the left and (or) the right views from which the virtual view is generated. It can thus be considered a reduced-reference method since the original reference of the picture under evaluation is not available. Fig. 1 shows the block diagram of the proposed algorithm for two view based DIBR.

The proposed ‘DIBR-Synthesized image Quality Metric (DSQM)’ is a block-based algorithm that works in three steps. In the first step, it divides the left and right view images into smaller blocks. The matching blocks in the synthesized image are found either by using the disparity information or detected through registration techniques in case the disparity information is unavailable. In the second step, the perceptual features of the two corresponding blocks are calculated. In the final step, these perceptual features are compared to estimate the distortion in the synthesized blocks; the overall quality of the synthesized image is estimated averaging the results obtained on all blocks.

3.1. Image block matching

A typical DIBR algorithm uses one or two camera views with corresponding depth maps to estimate the image at a virtual viewpoint. To keep the discussion simple, without the loss of generality we assume a 1D parallel camera setup where \( I_r \) and \( I_l \) are the original right and left camera views and \( I_v \) is a virtual view obtained by using DIBR which is subject to quality evaluation. Let the size of all the images be \( M \times N \).

To estimate the quality of the synthesized image, the proposed algorithm uses \( I_r \) and \( I_l \) as reference images. In the rest of the paper, we refer the left and the right view image as ‘input image’ and the DIBR image as the ‘synthesized image’. In the first step, the proposed algorithm divides the input images into small blocks of size \( m \times n \), and the corre-
perceptual features in images. The noisy and imperfect depth map distorts these perceptual features resulting in poor quality synthesized image. We use the Phase Congruency model [14, 27] to estimate the perceptual image features. Phase congruency model is a frequency based measure, according to which the locations where the Fourier components are maximally in phase represent the important perceptual features in the image. Phase congruency (PC) of a signal \( I(x) \) at point \( x \) is computed as [14]:

\[
PC(x) = \max_{\phi(n)\in[0,2\pi]} \frac{\sum_n A_n \cos(\phi_n(x) - \bar{\phi}(x))}{\sum_n A_n}
\]  

(3)

where \( A_n \) and \( \phi_n(x) \) represent the amplitude and the local phase of the \( n \)-th Fourier component at position \( x \) respectively. An efficient implementation of the above definition of PC is not viable. In [28] it is shown that phase congruency is equivalent to the ratio of energy and the sum of the Fourier amplitudes. That is:

\[
PC(x) = \frac{E(x)}{\sum_n A_n + \epsilon}
\]  

(4)

where \( E(x) = \sqrt{F^2(x) + H^2(x)} \) where \( F(x) \) is the signal with its DC component removed, \( H(x) \) is the Hilbert transform of \( F(x) \) and \( \epsilon \) is a small constant to keep the equation stable. Kovesi [27] proposed an efficient logarithmic Gabor wavelets based implementation of PC. Let \( e_n(x) \) and \( o_n(x) \) be the responses of the even-symmetric and odd-symmetric filters on scale \( n \). Kovesi showed that:

\[
F(x) \simeq \sum_n e_n(x), \quad H(x) \simeq \sum_n o_n(x),
\]

and

\[
\sum_n A_n \simeq \sum_n \sqrt{e_n(x)^2 + o_n(x)^2}.
\]

We use this logarithmic Gabor wavelets based PC to compute the perceptual features of the corresponding blocks \( p \) and \( q \). Moreover, from experiments we note that luminance based PC performs marginally better than RGB while assessing the synthesized image quality. Therefore, before computing the perceptual features, the RGB images are converted to YIQ color model. We use only the luminance component of the blocks to compute the perceptual features. Let \( PC_p \) and \( PC_q \) be the phase congruency transforms of luminance component of the blocks \( p \) and \( q \) respectively. Fig. 3 shows sample corresponding blocks \( p \) and \( q \) and their PC transforms. The mean of the phase congruency image is taken as perceptual feature.
3.3. Computing the quality of the synthesized image

The absolute difference $Q$ between the means of the phase congruency maps of the two corresponding blocks is computed to estimate the synthesis distortion due to DIBR:

$$Q = |\mu(\text{PC}_p) - \mu(\text{PC}_q)|$$  \hspace{1cm} (5)

where $\mu(\cdot)$ is the mean of the phase congruency map. The overall quality of the synthesized image is computed by averaging the quality scores of the matching block pairs. Let $\kappa$ be the number of the total matching block pairs. The overall quality of the synthesized image is computed as:

$$DSQM = \frac{1}{\kappa} \sum_{i=1}^{\kappa} Q_i$$ \hspace{1cm} (6)

where $Q_i$ is the quality of the $i$-th matching pair. DSQM is a distortion measure, i.e. smaller values represent better image quality.

4. EXPERIMENTAL EVALUATION AND RESULTS

4.1. Dataset used for experimental evaluation

The performance of the proposed quality metric is evaluated on DIBR-based synthesized images. We used the IRCCyN/IVC DIBR Images database\(^1\) [12] which is built from three standard MVD sequences: Book Arrival (16 cameras with 6.5 cm spacing), Lovebird1 (12 cameras with 3.5 cm spacing) and Newspaper (9 cameras with 5 cm spacing). Further details are presented in Tab. 1. From each sequence four different viewpoint images were obtained using the following seven well-known DIBR algorithms:

- A1: based on [4] where the depth map is preprocessed by using a low-pass filter to remove depth discontinuities. The borders of the synthesized image are cropped and image is resized to the original size.
- A2: based on [4] where holes are estimated using inpainting algorithm proposed in [29].
- A3: Tanimoto et al. [7], it is the View Synthesis Reference Software (VSRS).
- A4: Müller et al. [5] exploits the depth information to fill the holes.
- A5: Ndjiki-Nya et al. [8], it uses patch-based texture synthesis to estimate the holes.
- A6: Koppel et al. [30] uses the video temporal information and constructs a background sprite which is used to fill the holes.
- A7: refers to the synthesized sequences with unfilled holes.

Hence, for each sequence 28 synthesized video sequences were obtained and 84 sequences in total were generated. From each synthesized view sequence, only one frame was selected for subjective evaluation, hence the database consists of 84 still images which were rated for quality by 48 subjects using absolute category rating (ACR) as test methodology to obtain the mean opinion score (MOS). The references of the synthesized images were also rated and were used to obtain the differential mean opinion score (DMOS).

4.2. Performance evaluation results

The performance of the proposed DSQM metric is evaluated by comparing the objective scores with the subjective ratings. The performance is evaluated using the Pearson linear correlation coefficient (PLCC) for prediction accuracy test, the Spearman rank order correlation coefficient (SROCC) for prediction monotonicity test, and root mean square error (RMSE) is used to compute the prediction error. Before computing the performance parameters, as recommended by Video Quality Expert Group (VQEG) [31] the objective scores are mapped to the subjective ratings (DMOS) with monotonic nonlinear regression. The logistic function outlined in [32] is used for regression mapping:

$$DMOS_p = \beta_1 \left( \frac{1}{2} - \frac{1}{1 + \exp \beta_2(S - \beta_3)} \right) + \beta_4 S + \beta_5$$

where $S$ is the objective score, $DMOS_p$ is the mapped score and $\beta_1, \ldots, \beta_5$ are the regression model parameters. In all experiments, the maximum disparity parameter ($D$) is 32. The block size ($m \times n$) is empirically computed and set to $128 \times 128$.

To evaluate the effectiveness of the proposed algorithm, we compare its performance with various 2D and 3D IQA algorithms. In 2D-IQA algorithms, we chose the well-known

\begin{table}[h]
\begin{center}
\begin{tabular}{|l|l|l|}
\hline
Sequence & Size & Frame Number & Viewpoint \\
\hline
Book_Arrival & 1024\times768 & Left: frame 54 of the view 8 & Target center point of view: view 9 \\
 & & Right: frame 60 of the view 10 & \\
Lovebird1 & 1024\times768 & Left: frame 104 of the view 4 & Target center point of view: view 7 \\
 & & Right: frame 112 of the view 8 & \\
Newspaper & 1024\times768 & Left: frame 136 of the view 4 & Target center point of view: view 5 \\
 & & Right: frame 104 of the view 6 & \\
\hline
\end{tabular}
\end{center}
\caption{IRCCyN/IVC DIBR Images database details.}
\end{table}

\(^1\)http://ivc.univ-nantes.fr/en/databases/DIBR_Images/
Table 2. Overall performance of DSQM and other 3D-IQA algorithms on IRCCyN/IVC DIBR Images database.

<table>
<thead>
<tr>
<th>Metric</th>
<th>3DSwIM</th>
<th>Tsai</th>
<th>PQM</th>
<th>Gorley</th>
<th>SIQE</th>
<th>Dragana</th>
<th>MW-PSNR</th>
<th>DSQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLCC</td>
<td>0.6420</td>
<td>0.4830</td>
<td>0.4709</td>
<td>0.3183</td>
<td>0.6058</td>
<td>0.7191</td>
<td>0.7378</td>
<td><strong>0.7895</strong></td>
</tr>
<tr>
<td>SROCC</td>
<td>0.5613</td>
<td>0.4740</td>
<td>0.4869</td>
<td>0.3000</td>
<td>0.4347</td>
<td>0.6846</td>
<td>0.7070</td>
<td><strong>0.7151</strong></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.5105</td>
<td>0.5975</td>
<td>0.5877</td>
<td>0.6312</td>
<td>0.5298</td>
<td>0.4627</td>
<td>0.4494</td>
<td><strong>0.4086</strong></td>
</tr>
</tbody>
</table>

Table 3. Overall performance of DSQM and other 2D-IQA algorithms on IRCCyN/IVC DIBR Images database.

<table>
<thead>
<tr>
<th>Metric</th>
<th>SSIM</th>
<th>MS-SSIM</th>
<th>PSNR</th>
<th>VSNR</th>
<th>DSQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLCC</td>
<td>0.5639</td>
<td>0.5489</td>
<td>0.4283</td>
<td>0.5145</td>
<td><strong>0.7895</strong></td>
</tr>
<tr>
<td>SROCC</td>
<td>0.4687</td>
<td>0.5324</td>
<td>0.4628</td>
<td>0.5051</td>
<td><strong>0.7151</strong></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.5499</td>
<td>0.5566</td>
<td>0.6017</td>
<td>0.5709</td>
<td><strong>0.4086</strong></td>
</tr>
</tbody>
</table>

SSIM [33], MSSIM [34], PSNR, and VSNR [35] for performance comparison. Before computing the performance parameters, the objective scores predicted by these metrics are mapped to subjective ratings using the same logistic function (Eq. 7). The results are presented in Tab. 3 and clearly show that the proposed DSQM outperforms all 2D-IQA algorithms by a significant extent.

Now, we compare the performance of the proposed algorithm with the well-known 3D image quality assessment algorithms, namely Tsai [18], PQM [16], 3DSwIM [20], Gorley [21], SIQE [22], Dragana [26], and MW-PSNR [25]. The obtained correlation values are reported in Tab. 2, which show that the proposed algorithm achieves the best correlation with the subjective ratings and outperforms the competing 3D-IQA algorithms. In particular, it can be noted that DSQM achieves the correlation scores: 0.7895 in terms of PLCC and 0.7151 in terms of SROCC with the minimum RMSE (0.4086). The results presented in Tab. 3 and 2 demonstrate that the performance of proposed algorithm is appreciably better than other 3D and 2D quality assessment algorithms in evaluating the quality of DIBR synthesized images. Further details on performance evaluation of DSQM and comparison with other 3D-IQA algorithms, and the software to reproduce the results are available at http://www.di.unito.it/~farid/3DQA/DSQM.html.

5. CONCLUSIONS

In this paper, we proposed a quality assessment algorithm to estimate the quality of the DIBR synthesized images. The algorithm uses the left and right views to assess the synthesis distortion in the virtual image. It is a block-based algorithm that extracts the phase congruency based perceptual features from the input views (left and right view images) and the virtual image, and compares them to estimate the quality of the DIBR synthesized image. The experimental results demonstrate that the proposed algorithm performs better than other 2D and 3D-IQA algorithms.

6. REFERENCES


