



**HAL**  
open science

# A benchmark of DIBR Synthesized View Quality Assessment Metrics on a new database for Immersive Media Applications

Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges

► **To cite this version:**

Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges. A benchmark of DIBR Synthesized View Quality Assessment Metrics on a new database for Immersive Media Applications. *IEEE Transactions on Multimedia*, 2019, 21 (5), pp.1235-1247. 10.1109/TMM.2018.2875307 . hal-01899625

**HAL Id: hal-01899625**

**<https://hal.science/hal-01899625>**

Submitted on 7 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A benchmark of DIBR Synthesized View Quality Assessment Metrics on a new database for Immersive Media Applications

Shishun Tian, *Student Member, IEEE*, Lu Zhang, Luce Morin, and Olivier Déforges

**Abstract**—Depth-Image-Based Rendering (DIBR) is a fundamental technology in several 3D-related applications, such as Free viewpoint video (FVV), Virtual Reality (VR) and Augmented Reality (AR). However, new challenges have also been brought in assessing the quality of DIBR-synthesized views since this process induces some new types of distortions, which are inherently different from the distortion caused by video coding. In this paper, we present a new DIBR-synthesized image database with the associated subjective scores. We also test the performances of the state-of-the-art objective quality metrics on this database. This work focuses on the distortions only induced by different DIBR synthesis methods. Seven state-of-the-art DIBR algorithms, including inter-view synthesis and single view based synthesis methods, are considered in this database. The quality of synthesized views was assessed subjectively by 41 observers and objectively using 14 state-of-the-art objective metrics. Subjective test results show that the interview synthesis methods, having more input information, significantly outperform the single view based ones. Correlation results between the tested objective metrics and the subjective scores on this database reveal that further studies are still needed for a better objective quality metric dedicated to the DIBR-synthesized views.

**Index Terms**—Depth-Image-Based-Rendering (DIBR), FVV, view synthesis, QoE, quality assessment.

## I. INTRODUCTION

NOWADAYS, 3D video applications, such as 3D-TV and Free-viewpoint TV (FTV) [1], have received tremendous public attention because they can create a more realistic and immersive viewing experience. Especially, FTV allows the users to view a 3D scene by freely changing their viewpoints. For example, Canon announced on September 2017 its Free Viewpoint Video System that gives the users a better quality of experience (QoE) where they can view sporting events from various different angles and viewpoints, cf. Fig. 1.

As one widely accepted data representation format for 3D scenes, the Multiview Video plus Depth (MVD) format [2] consists of multiple texture images and their corresponding depth maps at some particular viewpoints. The other views are then synthesized through the Depth-Image-Based-Rendering (DIBR) technique [3]. The idea of DIBR is to synthesize the virtual views by using the texture and depth information at another viewpoint. Firstly, the image points at the original

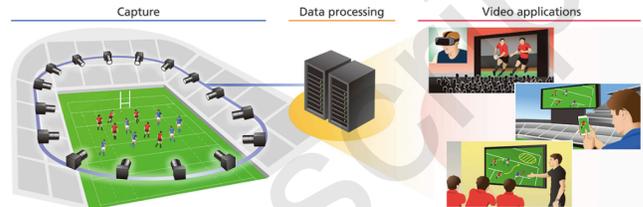


Fig. 1: Example of Canon's Free Viewpoint Video System. This figure is from <http://global.canon/en/news/2017/20170921.html>

viewpoint are reprojected into the 3D world by using its associated depth data. Then, these 3D points are projected into the image plane at the virtual target viewpoint. With the MVD format and the DIBR technology, only a limited number of original views and their corresponding depth maps are needed to be captured, stored and transmitted. In this way, the 3D multiview data compression and transmission can use the existing framework.

Great efforts have been put into investigating DIBR technology recently [5], [6], [7], [8], [9], DIBR is not only useful in the FTV [10], but also a promising solution for synthesizing virtual views in many other recent popular immersive multimedia applications, such as Virtual Reality (VR) [11], Augmented Reality (AR) [12] and light field multiview videos [13], etc. For example, the DIBR has already been used in a light field compression scheme where only very sparse samples (four views at the corners) of light field views are transmitted while the others are synthesized (cf. Fig. 2). This new scheme significantly outperformed HEVC inter coding for the tested light field images [4]. Another example concerns 360-degree and volumetric videos: two developing areas pointing to how video will evolve as VR/AR technology becomes the mainstream [14]. Current 360-degree videos allow viewers to look around in all directions, but only at the shooting location. They do not take into account the translation (changes in position) of the head. To make the QoE more immersive, some companies propose to use DIBR to synthesize the non-captured views when users move from the physical camera's position, as proposed in Technicolor's volumetric video streaming demonstration [15]. One similar approach can be found in Fig. 3, where typical DIBR artifacts appear (around the contours) if users try to look at an object on the floor hidden behind the person. In the social and embodiment VR media applications, where a VR media

This paper is supported by China Scholarship Council.

S. Tian, L. Zhang, L. Morin and O. Déforges are with Univ. Rennes, National Institute of Applied Sciences (INSA Rennes) and IETR UMR CNRS 6164, Rennes 35708 Cedex 7, France (e-mail: shishun.tian@insa-rennes.fr; lu.ge@insa-rennes.fr; luce.morin@insa-rennes.fr; olivier.deforges@insa-rennes.fr)



(a) A light field image captured by a camera at a corner



(b) A light field image synthesized by the DIBR

Fig. 2: Example of light field image rendered by DIBR in [4]

Fig. 3: Augmented Reality Screen Capture. This figure is from <https://developer.att.com/blog/shape-future-of-video>

designed for 360-degree videos mixed with real-time objects for multiple users, an eye-contact technique based on the DIBR [15] can provide the users the viewpoint according to their eye positions, which gives the users a better interactive QoE.

Although DIBR has a great potential to give users a better QoE in many immersive multimedia broadcasting applications, current DIBR algorithms may introduce some new types of distortions which are quite different from those caused by image compression. Most compression methods can cause specific distortions [16], eg. blur [17], blockiness [18] and ringing [19]. These distortions are often scattered over the whole image, while the DIBR-synthesized artifacts (caused by distorted depth map and imperfect view synthesis method) mostly occur in the disoccluded areas. Since most of the commonly used 2D objective quality metrics are initially designed to assess common coding distortions, they may fail in assessing the quality of DIBR-synthesized images [20], [21]. Towards new and better quality metrics for 3D synthesized view assessment, several publicly accessible databases have been developed, as shown in Table I.

Bosc et al. proposed the IRCCyN/IVC DIBR image database [22], [23] and IRCCyN/IVC DIBR video database [24]. The source content of these two database is extracted from 3 different MVD sequences: *BookArrival*, *Lovebird1*

and *Newspaper*. For each sequence, four virtual views are synthesized by seven DIBR view synthesis algorithms [3], [25], [26], [27], [28], [29], [30]. That is to say, there are 84 synthesized images or videos in each database. Besides, the IVC DIBR video database also provided three H.264 coded videos with three different quantization levels for each sequence. For each synthesized virtual image or video, the image/video captured by a real camera on the same viewpoint is used as reference. One big issue is that several DIBR algorithms tested in this database introduce some “old-fashioned” artifacts (such as “black holes”) which no longer exist when the state-of-the-art DIBR algorithms are used.

Song et al. proposed a publicly accessible stereoscopic 3D Database (MCL-3D database) for the quality assessment of DIBR-synthesized stereoscopic images in [31]. The DIBR technology is used to generate the left and the right views by using the 2D-image-plus-depth source. Many types of distortions are considered in this database, such as Gaussian blur, additive white noise, down-sampling blur, JPEG and JPEG-2000 (JP2K) compression and transmission error. These distortions are applied on either the original texture images or the depth images before the view synthesis. In addition, the distortion caused by imperfect DIBR algorithms are also considered in this database. Nine MVD sequences are collected, among which *Kendo*, *Lovebird1*, *Balloons*, *PoznanStreet* and *PoznanHall2* are natural images; *Shark*, *Microworld*, *GTFly* and *Undodancer* are Computer Graphics images. Four DIBR view synthesis algorithms ([3], [25], [32] plus DIBR without hole filling) were used. This database contains various types of distortions, but distortion types directly related to DIBR algorithms are very limited. The tested DIBR algorithms also produce some “old-fashioned” artifacts.

The SIAT Synthesized Video Quality Database [33] proposed by Liu et al. focused on the distortions introduced by compressed texture and depth images. For each of the ten different MVD sequences, 14 different texture/depth quantization combinations were used to generate the texture/depth view pairs with compression distortions. Then, the virtual videos are synthesized using the VSRS-1D-Fast software implemented in the 3D-HEVC [34] reference software HTM. Here, only compression distortions are evaluated.

TABLE I: Summary of existing DIBR related database

Name	No. seq.	No. DIBR algos	DIBR algos		other distortions	size	Reference	display	Observer
			Name	Year					
IVC DIBR-image	3	7	Fehn's	2004	No	84	original images from MVD seq.	2D	43
			Telea's	2003					
			VSRs	2009					
			Miller	2008					
			Ndjiki-Nya	2010					
			Kppel	2010					
Black hole	—								
IVC DIBR-video	3	7	idem		H.264	84	original images from MVD seq.	2D	32
MCL-3D	9	4	Fehn's	2004	Additive White Noise	693	images synthesized by VSRs	Stereo.	270
			Telea's	2003	Blur				
			HHF	2012	Down sampling				
			Black hole	—	JPEG				
					JPEG2k				
					Translation Loss				
SIAT video	10	1	VSRs	2009	3DV-ATM coding	140	original videos from MVD seq.	2D	56
IVY	7	4	Criminisi	2004	No	84	original images from MVD seq.	Stereo.	20
			Ahn's	2013					
			VSRs	2009					
			Yoon	2014					
Proposed	10	7	Criminisi	2004	No	140	original images from MVD seq.	2D	42
			VSRs	2009					
			LDI	2011					
			HHF	2012					
			Ahn's	2013					
			Luo's	2016					
Zhu's	2016								

Later, Jung et al. proposed another IVY stereoscopic 3D image database to assess the quality of DIBR synthesized stereoscopic images [35]. A total of 7 sequences are selected from four Middlebury datasets [36] (*Aloe*, *Dolls*, *Reindeer*, and *Laundry*) and three MVD sequences (*Lovebird1*, *Newspaper* and *Bookarrival*). 84 stereo image pairs are synthesized by four DIBR algorithms [37], [38], [39], [40] in this database. Note that in this database, virtual views were only generated by view extrapolation.

In this work, the proposed image database focuses on the distortions only caused by DIBR algorithms (like the IRCCyN/IVC DIBR database), but with state-of-the-art DIBR algorithms. The “view” or stimuli in this subjective test indicates an individual synthesized image. In total, we tested seven DIBR algorithms, including both the interview synthesis and the single view synthesis methods. We selected those DIBR algorithms which produce no longer “old-fashioned” artifacts and of which the code sources were provided by their authors. Note that the SIAT database focuses on the effect of texture and depth compression on the synthesized views and it contains only one DIBR algorithm. Compared to the MCL-3D and the IVY databases, the proposed new database (1) includes not only virtual views generated by view extrapolation, but also by view interpolation; (2) tests more and newer DIBR algorithms; (3) shows the views on a 2D display to avoid the 3D display settings and configurations influences (same approach was used in the IRCCyN/IVC DIBR database). The IRCCyN/IVC DIBR database also focuses on the comparison of different DIBR algorithms, but it contains some “old-fashioned” DIBR artifacts (eg. black holes) and it contains less source images than ours. The proposed database can be used along with the IRCCyN/IVC database for this type of usage. The main contributions of this paper are: (1) a new

publicly accessible DIBR synthesized image quality database with more recent DIBR algorithms; (2) a relatively complete bench marking of the state-of-the-art objective metrics for DIBR synthesized image quality assessment.

The rest of this paper is organized as follows. Section II briefly introduces the seven DIBR algorithms used in this database. The subjective experiments and the objective metrics performance study are described in Section III and Section IV respectively. In the end, the conclusions are drawn in Section V.

## II. TESTED DIBR ALGORITHMS

In this section, the tested DIBR algorithms are introduced. The DIBR is a process of generating novel views of a scene from original texture images and associated depth information, which can be divided into two steps: firstly, the original texture image is re-projected into 3D world aided by the associated per-pixel depth data; then these 3D space points are projected into the image plane of a new virtual view position. This concatenation of 2D-to-3D reprojection and the subsequent 3D-to-2D projection is usually called 3D image warping in the Computer Graphics (CG) literature. In this work, both inter-view interpolation and single view based synthesis methods are taken into consideration. The interview DIBR algorithm uses the two neighboring views to synthesize the virtual viewpoint, while the single view based synthesis methods only use one neighboring view to extrapolate the synthesized view.

The major problem of the DIBR method is the disocclusion: regions which are occluded in the captured views become visible in the virtual ones. Due to the lack of original texture information, a synthesized image often contains disocclusion holes which significantly degrades the quality. The processing of these disocclusion holes plays an important role in generating a synthesized view of high quality. The method to fill

them is known as the *disocclusion inpainting*, which is a special case of *image inpainting*. Thus, inpainting methods are part of the DIBR algorithms.

#### A. Criminisi's Exemplar based inpainting

Criminisi et al. proposed a new algorithm for *image inpainting*. As shown in Fig. 4, it employs an exemplar-based texture synthesis technique [37]. A *confidence* is used to compute patch priorities, and to optimize the fill order of the target regions according to their priorities. The actual color values are computed using exemplar-based synthesis. After the target patch has been filled with new values, the *confidence* in this patch is updated. The *confidence* in the synthesized pixel values is propagated in a manner similar to the propagation of information in inpainting. As filling proceeds, confidence values decay, which indicates that the pixel color values are less reliable near the center of the target region.

#### B. LDI

Jantet et al. proposed an object-based Layered Depth Image (LDI) representation to improve the quality of virtual synthesized views [41]. As shown in Fig. 5, they firstly segment the foreground and background based on a region growing algorithm, which allows organising LDI pixels into two object-based layers. Once the extracted foreground is obtained, an inpainting method is used to reconstruct the complete background image on both depth and texture images. Several inpainting method can be chosen, for example, Navier-Stoke based method [42], Telea method [25] and Gautier method [43]. In this work, the Gautier inpainting method is used in the LDI.

#### C. Ahn's method

Ahn et al. proposed a depth based disocclusion filling method using patch-based texture synthesis [38]. Firstly, a median filtering is applied to texture and depth images to remove the small cracks caused by rounding errors in the 3D warping process. In order to handle the ghost effect due to mismatch of the boundaries of the foreground objects in the texture and depth image, a ghost effect removal method is added in the 3D warping process. During the *disocclusion inpainting* procedure, the Criminisi's method is improved by optimizing the filling priority and the patch-matching measure. The new priority term uses the Hessian matrix structure tensor which is robust to noise and reflects the overall structure of an image area. The optimized matched patch is selected through the data term on the background regions which were extracted using warped depth map. The filling of disoccluded holes in a depth map is conducted simultaneously with filling holes in the texture image. The block diagram of Ahn's view synthesis method is shown in Fig. 6.

#### D. Luo's method

Luo et al proposed a hole filling approach for DIBR systems based on background reconstruction [44]. As shown in Fig. 7, in order to extract the foreground, the depth map is firstly

preprocessed by a cross-bilateral filter and morphological operations. Then the Canny's edge detection is employed to extract the initial seeds for random walker, and the foreground is finally extracted from the depth map by random walker segmentation. After the removal of foreground, the temporal correlation information in both the 2D video and its corresponding depth map is exploited to construct a background video based on motion compensation and modified Gaussian Mixture model. Finally, the reconstructed background video is warped to the virtual viewpoint to eliminate the disocclusion holes.

#### E. HHF-Hierarchical hole-filling

Solh et al. proposed two pyramid-like approaches, namely Hierarchical Hole-Filling (HHF) and Depth Adaptive Hierarchical Hole-Filling, to eliminate the disoccluded holes in DIBR synthesized views [32]. The block diagram of HHF is shown in Fig. 8, which can be divided into four steps. Firstly, a sequence of images  $R_0, \dots, R_N$  are low-pass filtered using a pseudo Gaussian plus zero elimination filtering operation (*reduce*), in which the original 3D warped image is marked as  $R_0$ .  $R_1$  is the *reduced* version of  $R_0$ , and so on. The Gaussian pyramid is generated by this *reduce* operation when the holes do not influence the calculations. Secondly, they start from the highest level of this pyramid  $R_N$ , an *Expand* operation is utilized to get an interpolated image  $E_{N-1}$ , whose size is equal to  $R_{N-1}$ . Then, this interpolated image  $E_{N-1}$  is used to fill the disoccluded holes in  $R_{N-1}$  to obtain the filled image  $F_{N-1}$ . Finally, the filled image in each scale,  $F_{N-1}, \dots, F_0$  can be obtained by repeating the operations upon, and  $F_0$  is the final inpainted result. The DAHHF method adds a depth adaptive preprocessing before the *reduce* and *expand* operations. Since the disoccluded regions are more likely to be the background regions, a depth map is employed to assign higher weights to the pixels belonging to the background. The following steps are similar to HHF except that the starting image is the preprocessed image and the depth weight must be considered during the *Fill* operations.

#### F. VSRS-View Synthesis Reference Software

Tanimoto et al. proposed a DIBR method [26] which has been adopted by the MPEG 3D video Group, known as View Synthesis Reference Software (VSRS) [39]. The depth discontinuity artifacts are solved by performing a post-filter on the projected depth map. Then, the inpainting method proposed in [25] is used to fill the holes in the disoccluded regions. This approach is primarily used in the inter-view synthesis applications which have just small holes to be filled, but it can also be used in single view based rendering cases. In this paper, both the interview mode (VSRS2) and the single view based mode (VSRS1) are used.

#### G. Zhu's method

Zhu et al. proposed a novel depth-enhanced hole filling approach for DIBR view interpolation [45]. Instead of inpainting the warped images directly, they focus on the use of

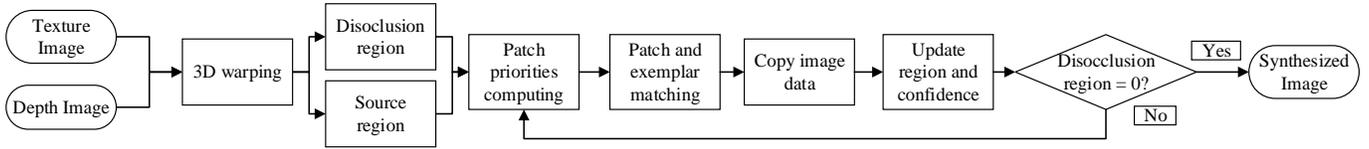


Fig. 4: Block diagram of Criminisi

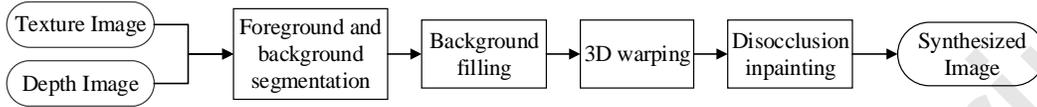


Fig. 5: Block diagram of LDI

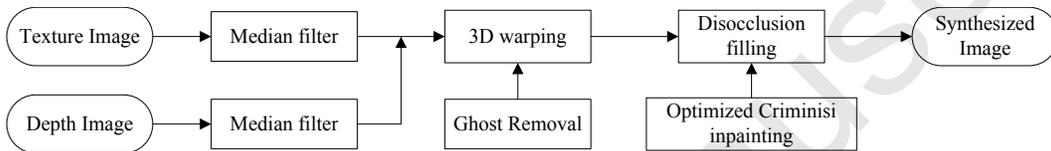


Fig. 6: Block diagram of Ahn's method

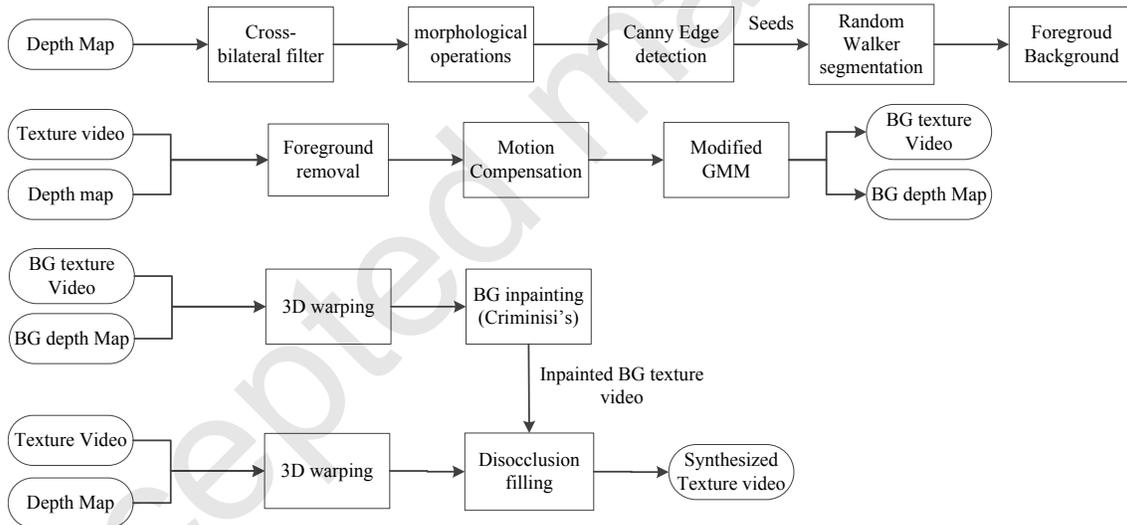


Fig. 7: Block diagram of Luo's method

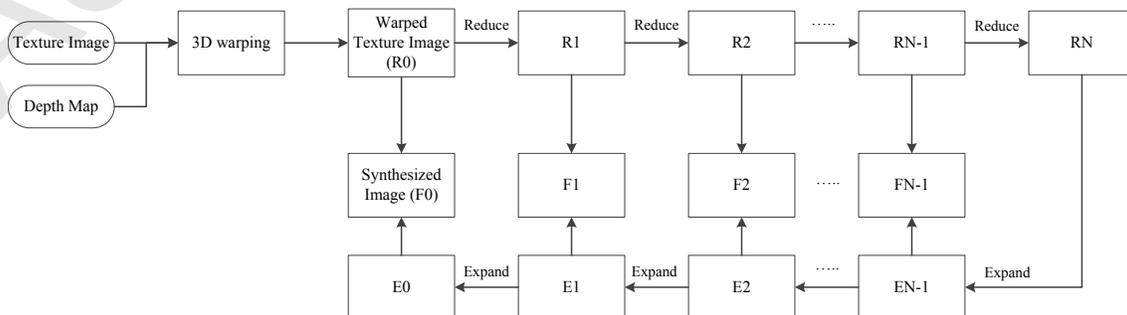


Fig. 8: Block diagram of HHH

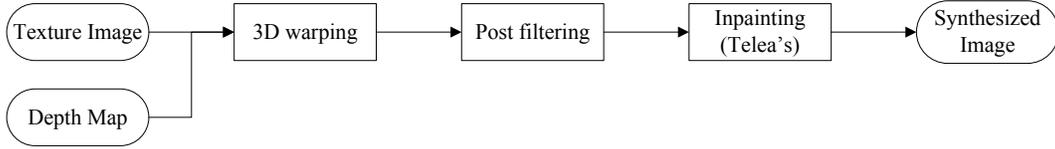


Fig. 9: Block diagram of VSRS

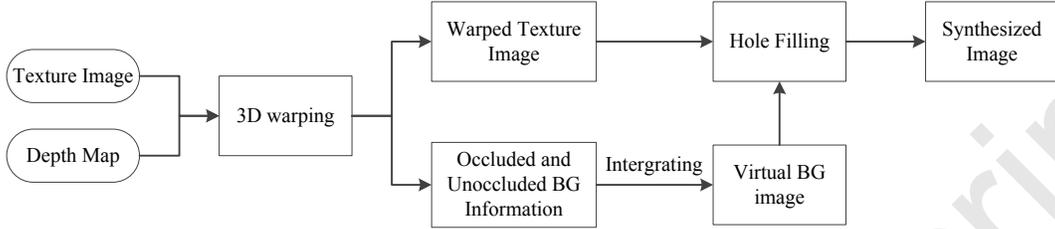


Fig. 10: Block diagram of Zhu's method

the occluded information to identify the relevant background pixels around the holes. Firstly, the occluded background information is registered in both texture and depth during the 3D warping process, and the background pixels around the holes are found. Then, the unoccluded background information around the holes is extracted based on the depth map. After that, a virtual image is generated by integrating the occluded background and unoccluded background information. The disoccluded holes are filled based on this generated image with the help of a depth-enhanced Criminisi's inpainting method and a simplified block-averaged filling method. Finally, the pre-stored foreground information is recovered in the virtual synthesized image.

Among the DIBR algorithms mentioned above, Zhu's method is an interview synthesis method, VSRS is used both as interview synthesis and single view based synthesis (marked as VSRS2 and VSRS1 respectively in this paper), the others are only single view based synthesis methods.

### III. SUBJECTIVE EXPERIMENT

Ten MVD test sequences provided by MPEG for the 3D video coding are used in this experiment. The *Balloons*, *BookArrival*, *Kendo*, *Lovebird1*, *Newspaper*, *Poznan Street* and *PoznanHall* sequences are natural images while the *Undo Dancer*, *Shark* and *Gt Fly* are computer animation images, as shown in Fig. 11. The characteristics of the sequences are summarized in Table II.

For each single view based DIBR algorithm, a single virtual viewpoint is extrapolated from the neighboring two views separately. For the interview DIBR algorithms, the virtual viewpoint is synthesized based on both the two neighboring views, as shown in Table III. We consider thus for each reference image, 2 virtual views synthesized by 2 interview synthesis algorithms and 12 virtual views synthesized by 6 single view based DIBR algorithm, which leads to 14 degraded images.

#### A. Subjective Test Methodology

There are several subjective testing methods to obtain the perceived quality scores, such as the subjective assessment

methodology for video quality (SAMVIQ) [46], the absolute categorical rating (ACR), etc. In this test, we choose to follow the SAMVIQ protocol because of its stability, reliability and relatively higher discriminability. The SAMVIQ results have a greater accuracy than the ACR scores for the same number of observers (on average 30% fewer observers were required for SAMVIQ than ACR for the same level of accuracy) [47].

In the SAMVIQ protocol, there is much more freedom for the observers who can view each image several times and correct the notation at any time they want. The observers can compare the degraded versions with each other, as well as with the explicit reference. In each trial, there is also a hidden reference which helps to evaluate the intrinsic quality of the reference when the perceived quality of the reference is not perfect. A continuous quality rating scale ranging from 0 to 100 is used during the test. It can be categorized according to the five quality levels: Bad, Poor, Fair, Good and Excellent. (See Table IV)

The experiment was conducted on a NEC MultiSync PA322UHD monitor with resolution  $3840 \times 2160$ . The environment of the subjective experiment was controlled as recommended in the ITU-R Rec. BT.1788 [48].

Altogether, 42 naive observers (28 males and 14 females with an age varying from 19 to 52 years old) participated in the subjective assessment experiment. All the observers have no prior knowledge of the view synthesis methodology domain. Prior to the test, the observers were screened for normal visual acuity on the Snellen chart, and for normal colour vision using the Ishihara chart. A training session was conducted before the test session. The observers could have a rest at any time they want during the test. The total duration of the experiment varied from 30 to 45 minutes for each observer.

#### B. Processing of Subjective Scores

The subjective scores were firstly processed using the observer screening method recommended in the ITU-R Rec. BT.1788 [48]. In this experiment, only one observer is eliminated after the observer screening. That leads to 41 observers finally for this database.

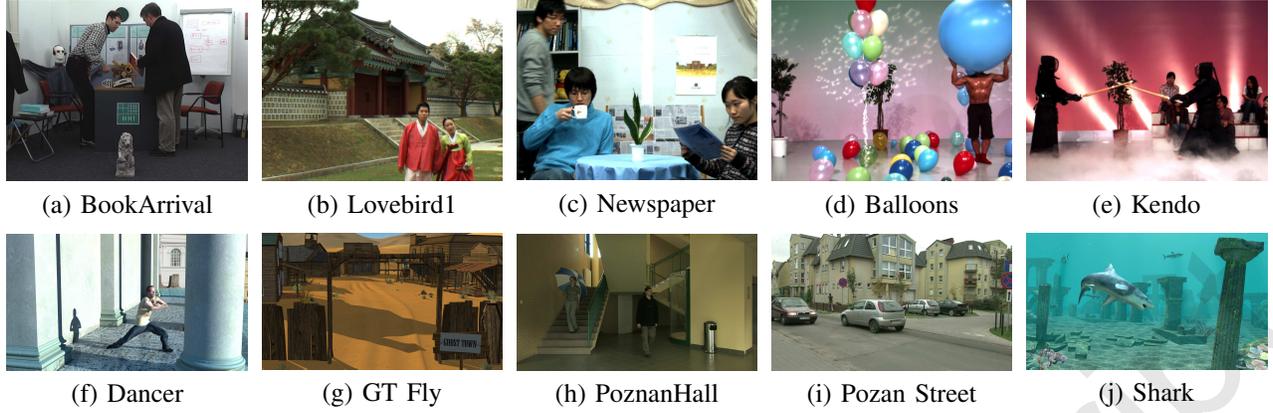


Fig. 11: The MVD sequences used in this paper

TABLE II: Introduction of the tested MVD sequences

Sequence	Resolution	Frame No.	View ref. Position	View sys. Position	SI
BookArrival	1024 × 768	58	8, 10	9	60.2348
Lovebird1	1024 × 768	80	4, 8	6	64.9756
Newspaper	1024 × 768	56	2, 6	4	61.1012
Balloons	1024 × 768	6	1, 5	3	47.6410
Kendo	1024 × 768	10	1, 5	3	48.6635
Undo Dancer	1920 × 1088	66	1, 9	5	64.1033
GT Fly	1920 × 1088	150	1, 9	5	55.5549
Poznan street	1920 × 1088	26	3, 5	4	61.3494
Poznan Hall2	1920 × 1088	150	5, 7	6	23.5174
Shark	1920 × 1088	220	1, 9	5	48.6635

TABLE III: Type of DIBR method

DIBR method	inter-view or single view (extrapolation)
VSRS2	inter-view
Zhu's	inter-view
Criminisi's	single view (extrapolation)
Luo's	single view (extrapolation)
HHF	single view (extrapolation)
LDI	single view (extrapolation)
VSRS1	single view (extrapolation)
Ahn's	single view (extrapolation)

TABLE IV: Comparison Scale for SAMVIQ

10	Bad
30	Poor
50	Fair
70	Good
90	Excellent

The primary quality scores of the tested image are obtained as the difference between the score of the hidden reference image and the score of the tested image as shown in Eq. (1):

$$S_{i,j} = Score_{hr,j} - Score_{i,j} \quad (1)$$

where  $S_{i,j}$  denotes the primary quality score of the  $i$ th tested synthesized image,  $Score_{hr}$  and  $Score_i$  denote the score of the hidden reference and the  $i$ th tested synthesized image respectively, and the subscript  $j$  denotes the  $j$ th observer.

Then, the primary quality scores are normalized to z-score per person cf. Eq. 2.

$$Zscore_{i,j} = \frac{S_{i,j} - \mu_j}{\sigma_j} \quad (2)$$

where  $\mu_j$  and  $\sigma_j$  denotes the mean value and variance value of the  $j$ th observer respectively. To make the data more intuitive, the normalized z-scores are scaled to (0,1).

The final quality score differential mean opinion score (DMOS) is calculated by averaging the normalized z-scores of all the observers, as shown in Eq. 3:

$$DMOS_i = \sum_{j=1}^N Zscore_{i,j} / N \quad (3)$$

where  $DMOS_i$  denotes the final subjective quality score of the  $i$ th tested synthesized image,  $S_{i,j}$  is the obtain primary quality score in Eq. (1), and  $N$  is the number of observers.

The obtained  $DMOS$  score distributions and their confidence intervals are shown in Fig. 12. Generally, the interview synthesis methods outperform the single view based synthesis methods in most sequences. However in some sequences, such as *BookArrival*, the VSRS1 get better results than VSRS2 and Zhu's methods, but not very significantly according to the corresponding confidence intervals. One reason could be that, owing to the inaccuracy of depth map, the same object in the two base views are rendered to different positions which results in a "ghost" effect in the synthesized view. However, this situation does not happen in single view based synthesis method VSRS1. As shown in Fig. 13, there exists a "ghost" effect of the "chat flow" on the board marked by red blocks in (c) and (d); but according to the synthesized content marked by red circles, the interview synthesis methods (c), (d) works better than the single view based methods (a), (b) in generating the object texture.

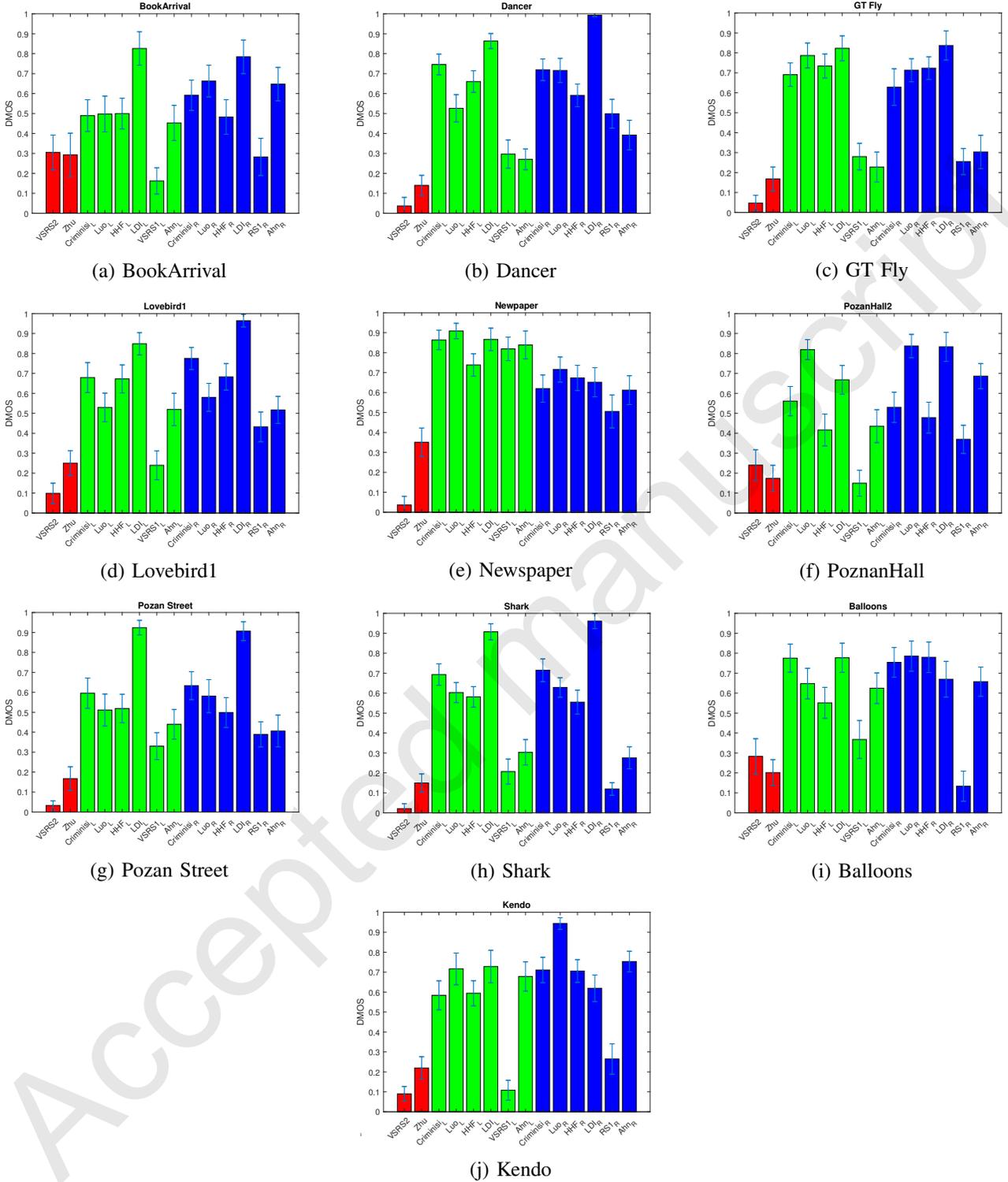


Fig. 12: DMOS distribution and confidence intervals of the synthesized views of different MVD sequences and different view synthesis methods. The x-labels are  $VSRS2$ ,  $Zhu$ ,  $Criminisi_L$ ,  $Luo_L$ ,  $HHF_L$ ,  $LDI_L$ ,  $VSRS1_L$ ,  $Ahn_L$ ,  $Criminisi_R$ ,  $Luo_R$ ,  $HHF_R$ ,  $LDI_R$ ,  $VSRS1_R$ ,  $Ahn_R$  ordinally. The subscript  $L$  means this virtual view is synthesized from the neighboring left view, while the subscript  $R$  means from the right.  $VSRS2$  denotes the view interpolation inter-view mode of VSRS. The error bars indicates the corresponding confidence intervals of the tested images. The bars referring to interview synthesize views are marked by red, the left-extrapolated views marked by green, and the right-extrapolated views marked by blue.

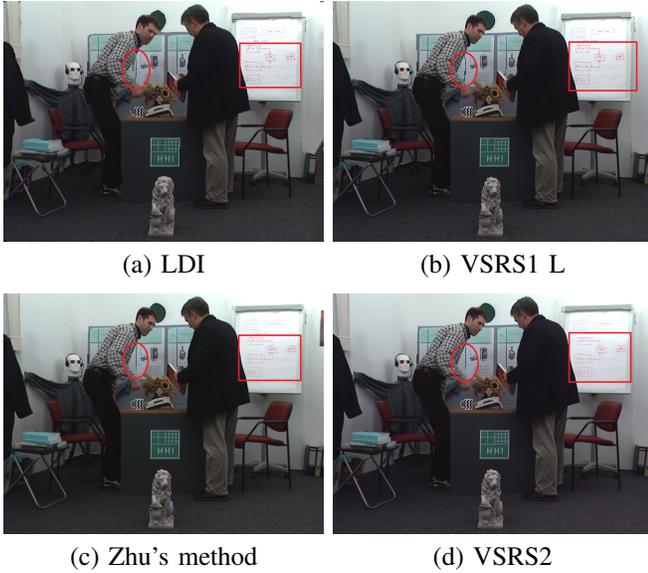


Fig. 13: Examples of synthesized images

A statistical analysis (student T-test here) was also made over the obtained *DMOS* scores, to show the statistical equivalence information of the tested algorithms. The scores of single view based methods are obtained by averaging the scores of the two images synthesized from the viewpoints at the two sides. As shown in Table V, the view interpolation methods (VSRS2 and Zhu's), which use the two neighboring views as reference views, perform much better than the single view based approaches. Among the single view based approaches, VSRS1 and Ahn's methods are significantly superior to the others.

TABLE V: Student T-test with obtained *DMOS* scores, where the symbol 1 indicates that the DIBR synthesis method in the row is significantly superior to the one in the column, the symbol -1 means the opposite, while 0 indicates that there is no significant difference between the DIBR synthesis methods in the row and in the column.

	VSRS2	Zhu	Cri.	Luo	HHF	LDI	VSRS1	Ahn
VSRS2	0	1	1	1	1	1	1	1
Zhu	—	0	1	1	1	1	1	1
Cri.	—	—	0	0	-1	1	-1	-1
Luo	—	—	—	0	-1	1	-1	-1
HHF	—	—	—	—	0	1	-1	-1
LDI	—	—	—	—	—	0	-1	-1
VSRS1	—	—	—	—	—	—	0	1
Ahn	—	—	—	—	—	—	—	0

#### IV. OBJECTIVE MEASUREMENT

In this section, we compare the performances of several existing objective image quality assessment metrics on the proposed database.

##### A. Objective Metrics

The tested image quality assessment (IQA) metrics are introduced below, where full-reference (FR) and no-reference (NR) are introduced respectively:

The full-reference (FR) 2D metrics include:

- SSIM: Structure SIMilarity, a widely used objective FR IQA metric calculating the structure similarity between the tested image and the reference image proposed by Wang et al. in [49].
- MS-SSIM: Multi-Scale Structure SIMilarity, a multi-scale approach of SSIM proposed by Wang et al. in [50].
- PSNR: Peak Signal to Noise Ratio, a widely used pixel-based metric.
- IW-PSNR, IW-SSIM: Information content Weighted FR IQA Metric based on PSNR and SSIM separately, proposed by Wang et al. in [51].
- UQI: Universal Quality Index proposed by Wang et al. in [52], models the image distortions by integrating loss correlation, luminance distortion and contrast distortion.
- PSNR-HVS: based on PSNR and UQI, takes the Human Vision System (HVS) into account [53], [54].

The FR 3D/synthesized-view-dedicated metrics include:

- MP-PSNR: Morphological Pyramid Peak Signal-to-Noise Ratio proposed by Sandic-Stankovic in [55], an FR IQA metric for DIBR-synthesized image based on multi-scale pyramid decomposition using morphological filters. The geometric distortions in different scales are obtained by comparing the differences between the reference and synthesized views in the corresponding scale.
- MW-PSNR: Morphological Wavelet Peak Signal-to-Noise Ratio by Sandic-Stankovic in [56], an FR IQA metric for DIBR-synthesized image. It is similar to the MP-PSNR except that the MW-PSNR uses morphological wavelet decomposition.
- MP-PSNR<sub>r</sub>, MW-PSNR<sub>r</sub>: the reduced version of MP-PSNR and MW-PSNR using only detail images from higher decomposition scales proposed by Sandic-Stankovic in [57].
- VSQA: View Synthesis Quality Assessment (VSQA) proposed by Conze et al. in [58], which uses three weighting maps to characterize the image complexity in terms of texture, diversity of gradient orientations and presence of high contrast.
- Tian's method: Tian et al. proposed a full-reference quality assessment metric for DIBR synthesized views [59], which uses an "affine" transform to compensate the global shift in the synthesized view and a disocclusion mask to weight the final distortions (considering that the synthesis distortion mainly occurs in the disoccluded regions).

The side view based full-reference (FR) metric:

- LOGS: Li et al. proposed a side view based FR metric for DIBR-synthesized views by measuring local geometric distortions in disoccluded regions and global sharpness (LOGS) in [60]. It uses the image at the original viewpoint instead of the image at the synthesized viewpoint as the reference.

The no-reference (NR) 2D metrics include:

- BIQI: Blind Image Quality Index, a NR IQA metric proposed by Moorthy et al. in [61].

- BliindSII: BLind Image Interfrity Notaor using DCT Statistics-II proposed by Saad et al. in [62].
- NIQE: Natural Image Quality Evaluator, a NR IQA metric proposed by Mittal et al. in [63].

The NR 3D/synthesized-view-dedicated metrics include:

- APT: Autoregression Plus Thresholding (APT) proposed by Gu et al. in [64]. They uses an autoregression based local image descriptor to detect the geometric distortions, followed by a saliency weighting and a thresholding to obtain the final quality measurement.
- NIQSV: An NR Image Quality metric for Synthesized views proposed by Tian et al. in [65]. They exploit simple opening and closing operations to remove the synthesis distortions, and then the quality is obtained through the comparison between the synthesized image and the processed synthesized image.
- NIQSV+: Improved NIQSV proposed by Tian et al. in [66].

### B. Correlation between the objective and subjective measurements

The performance of objective metrics can be evaluated by their correlation with the subjective quality scores. Normally, three criteria, i.e., the Pearson Linear Correlation Coefficients (PLCC), the Spearman's Rank Order Correlation Coefficients (SROCC) and the Root-Mean-Square-Error (RMSE), are used to measure the prediction accuracy, monotonicity and residuals respectively. Before the calculation of the three criteria above, the obtained objective scores need to be fitted to the predicted subjective score ( $DMOS_p$ ) by a non-linear regression function as suggested by the Video Quality Expert Group (VQEG) Phase I FR-TV [67]:

$$DMOS_p = a \cdot score^3 + b \cdot score^2 + c \cdot score + d \quad (4)$$

where  $score$  is the score obtained by the objective metric and  $a, b, c, d$  are the parameters of this cubic function. They are obtained through a regression to minimize the difference between  $DMOS_p$  and  $DMOS$ .

The obtained PLCC, RMSE, SROCC values are given in Table VI. It can be noticed at once that the performance of these metrics on the presented database is quite bad (no PLCC value more than 70%). Among which, Tian's FR 3D metric and the side view based FR metric LOGS perform the best in terms of the PLCC on this database. Especially for NIQSV+, NIQSV, NIQE and BliindS2 NR metrics, they show weak correlations with the subjective results, one reason may be that the NIQE and BliindS2 are trained and focus on the traditional artifacts, such as blurry, additive white noise, jpeg etc. they cannot well assess the quality of DIBR synthesized views. While the NIQSV and NIQSV+ metrics tried to optimize their performances on the IRCCyN/IVC DIBR database where "old-fashioned" artifacts exist, then cannot get a good performance when these artifacts are excluded.

In Table VII, we use the parameters provided by the authors, which make the algorithms achieve their best performance on IVC DIBR database. In Table VII, we investigate the performance dependency of MWPSNR and MPPSNR on

TABLE VI: PLCC, RMSE and SROCC between DMOS and objective metrics. Where "SV FR metric" indicates the side view based FR metric

Metric		PLCC	RMSE	SROCC
FR 2D metrics	PSNR	0.6012	0.1985	0.5356
	SSIM	0.4016	0.2275	0.2395
	MS-SSIM	0.6162	0.1957	0.5355
	IW-PSNR	0.5827	0.2019	0.4973
	IW-SSIM	0.6280	0.1933	0.5950
	UQI	0.4346	0.2237	0.4113
	PSNR-HVS	0.5982	0.1991	0.5195
FR 3D metrics	MP-PSNR	0.5753	0.2032	0.5507
	MP-PSNRr	0.6061	0.1976	0.5873
	MW-PSNR	0.5301	0.2106	0.4845
	MW-PSNRr	0.5403	0.2090	0.4946
	VSQA	0.5576	0.2062	0.4719
	Tian's	0.6685	0.1844	0.5903
SV FR metric	LOGS	0.6687	0.1845	0.6683
NR 3D metrics	NIQSV	0.1759	0.2446	0.1473
	NIQSV+	0.2095	0.2429	0.2190
	APT	0.4225	0.2252	0.4187
NR 2D metrics	NIQE	0.2244	0.2421	0.1360
	BliindS2	0.2225	0.2422	0.1329
	BIQI	0.4348	0.2237	0.4328

decomposition level and structural element size. It shows that these parameters can be fitted to achieve better performance on the proposed database, but they still cannot get satisfactory results. In addition, we think that a high degree of generality is a desirable feature for a good quality metric. That means a metrics performance cannot be judged only on its best performance on a selected database. This is also why the cross-validation is usually needed for the validation of a metric.

TABLE VII: Performance dependency of MP-PSNR and MW-PSNR, where SE indicates Structural Element size

		MW-PSNR			MW-PSNRr		
level	SE	PLCC	RMSE	SROCC	PLCC	RMSE	SROCC
8		0.5500	0.2070	0.5199	0.5602	0.2054	0.5235
7		0.5389	0.2088	0.4875	0.5383	0.2089	0.4953
6		0.6132	0.1958	0.5598	0.6095	0.1965	0.5634
5		0.5981	0.1987	0.5353	0.6014	0.1981	0.5240
level	SE	MP-PSNR			MP-PSNRr		
6	7	0.6037	0.1976	0.5578	0.5659	0.2044	0.5402
6	5	0.6284	0.1929	0.5737	0.5889	0.2004	0.5794
6	3	0.6312	0.1923	0.5914	0.6023	0.1979	0.5745
6	2	0.6134	0.1958	0.5601	0.5945	0.1993	0.5443
5	7	0.6190	0.1947	0.5809	0.5841	0.2012	0.5570
5	5	0.6294	0.1927	0.5951	0.6160	0.1953	0.5870
5	3	0.6246	0.1936	0.5860	0.6314	0.1922	0.5855
5	2	0.6163	0.1952	0.5497	0.6009	0.1982	0.5313
4	7	0.6170	0.1951	0.5909	0.6023	0.1979	0.5569
4	5	0.6230	0.1939	0.5796	0.6247	0.1936	0.5678
4	3	0.6170	0.1951	0.5619	0.6106	0.1963	0.5470
4	2	0.5911	0.2000	0.5319	0.5571	0.2059	0.4928

The scatter plot of each IQA metric is shown in Fig. 14. It seems that all methods are incapable of predicting worse qualities (bigger DMOS value indicates worse quality), which is however consistent with the results shown in Table VI where no metric has a PLCC value higher than 0.7. While some of the metrics do sometimes succeed in their prediction of high qualities in Fig. 14 (consistent with their PLCC values bigger than 0.5), the NR metrics NIQSV+, NIQSV, NIQE and BliindS2 show little correction with the subjective results (consistent with their PLCC values smaller than 0.3).

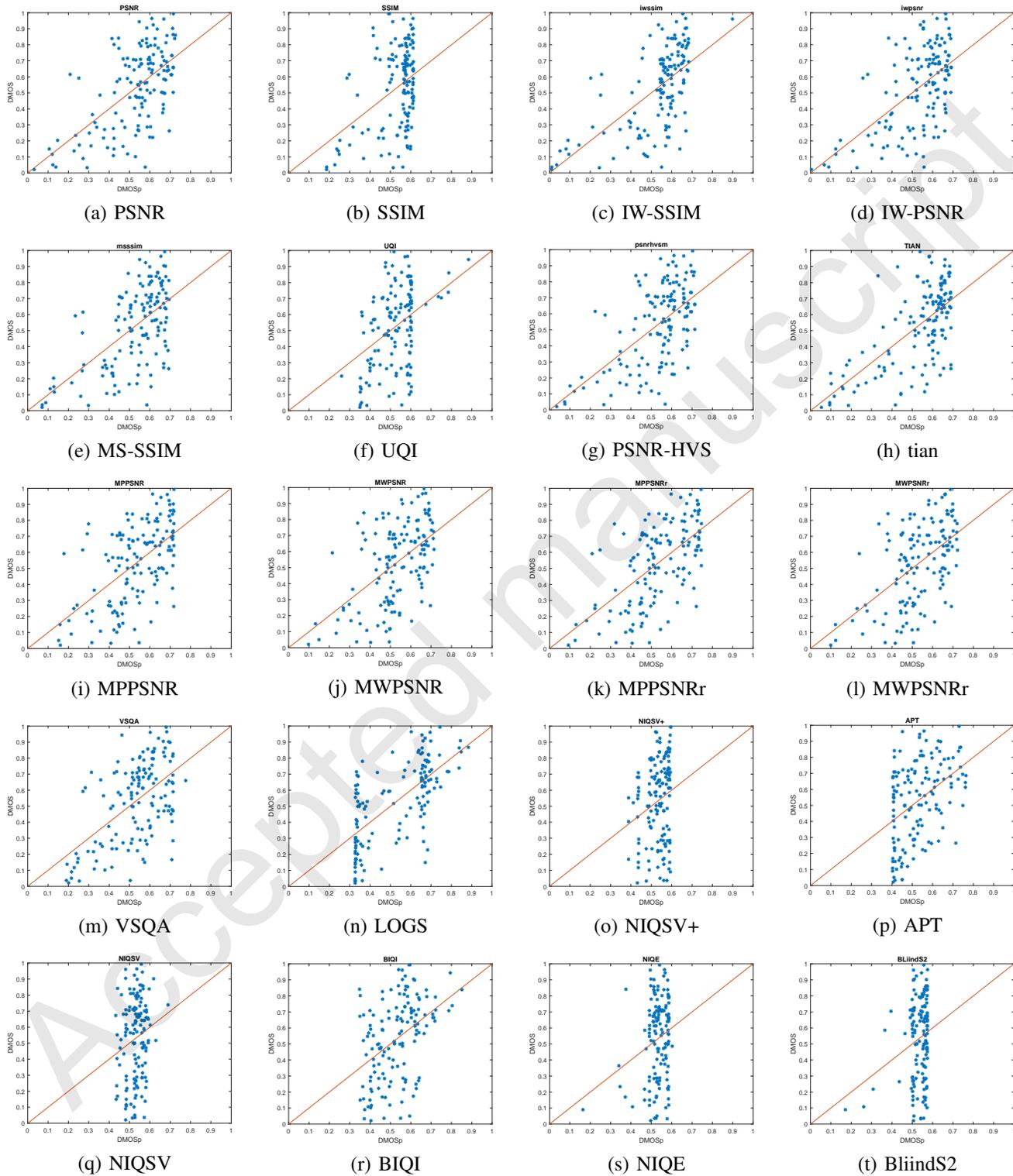


Fig. 14: Scatter plots of DMOS versus DMOSp of each IQA method

## V. CONCLUSION AND PERSPECTIVE

DIBR is widely used in FVV, VR, AR, and other popular topics considered as the next generation of 3D broadcasting applications, in order to provide a better QoE to users. In this paper, a new DIBR-synthesized image database which focuses on the distortions induced by different state-of-the-art DIBR view synthesis algorithms is presented. Ten MVD sequences and seven state-of-the-art DIBR view synthesis algorithms are selected to generate the virtual view images. The subjective experiment is conducted following the SAMVIQ protocol in a controlled environment as recommended by ITU-R Rec. BT.1788 [48]. Results show that the inter-view synthesis methods, which have more input information, significantly outperform the single view based synthesis algorithms. Furthermore, several objective measurements were used to assess the quality of synthesized images on this database. Their performance results indicate that further work has to be done to exploit deeply the characteristics of these specific distortions, for new objective metrics with a better correlation with subjective scores. In current database, only the MPEG MVD source images are included; in the future work, more source images, such as the images from Middlebury database [36], will be considered to make the experiment results more benchmark.

All the data of this presented database, including images, the ground truth depth maps and their associated DMOS, is publicly accessible ([https://vaader-data.insa-rennes.fr/data/stian/ieeetom/IETR\\_DIBR\\_Database.zip](https://vaader-data.insa-rennes.fr/data/stian/ieeetom/IETR_DIBR_Database.zip)), for the improvement of the QoE of DIBR related applications.

## ACKNOWLEDGMENT

The authors would like to thank China Scholarship Council for funding. The authors would also like to thank D.Sandić-Stanković [55], [56], [57], M. Solh [32], I. Ahn [38], J. Gautier [43], C. Zhu [45] and L. Li [60] for sharing their code or executable files.

## REFERENCES

- [1] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "Free-viewpoint TV," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 67–76, 2011.
- [2] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *2007 IEEE International Conference on Image Processing*, vol. 1, Sept 2007, pp. I – 201–I – 204.
- [3] C. Fehn, "Depth-Image-Based Rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 93–104.
- [4] X. Jiang, M. L. Pendu, and C. Guillemot, "Light field compression using depth image based view synthesis," in *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. IEEE, 2017, pp. 19–24.
- [5] S. Li, C. Zhu, and M.-T. Sun, "Hole filling with multiple reference views in dibr view synthesis," *IEEE Transactions on Multimedia*, 2018.
- [6] C. H. Cheung, K. N. Ngan, and L. Sheng, "Spatio-temporal disocclusion filling using novel sprite cells," *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1376–1391, 2018.
- [7] W.-N. Lie, C.-Y. Hsieh, and G.-S. Lin, "Key-frame-based background sprite generation for hole filling in depth image-based rendering," *IEEE Transactions on Multimedia*, vol. 20, no. 5, pp. 1075–1087, 2018.
- [8] M. Schmeing and X. Jiang, "Faithful disocclusion filling in depth image based rendering using superpixel-based inpainting," *IEEE Transactions on Multimedia*, vol. 17, no. 12, pp. 2160–2173, 2015.
- [9] Y. Mao, G. Cheung, and Y. Ji, "On constructing z-dimensional dibr-synthesized images," *IEEE Transactions on Multimedia*, vol. 18, no. 8, pp. 1453–1468, 2016.
- [10] D. Ren, S.-H. G. Chan, G. Cheung, V. Zhao, and P. Frossard, "Anchor view allocation for collaborative free viewpoint video streaming," *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 307–322, 2015.
- [11] H. Rheingold, *Virtual reality: exploring the brave new technologies*. Simon & Schuster Adult Publishing Group, 1991.
- [12] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and virtual environments*, vol. 6, no. 4, pp. 355–385, 1997.
- [13] N. Sabater, G. Boisson, B. Vandame, P. Kerbiriou, F. Babon, M. Hog, R. Gendrot, T. Langlois, O. Bureller, A. Schubert *et al.*, "Dataset and pipeline for multi-view light-field video," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017, pp. 1743–1753.
- [14] "The future of video: Enabling immersion," <https://developer.att.com/blog/shape-future-of-video>.
- [15] "Industrial demonstration," <https://igrv2017.sciencesconf.org/resource/page/id/12>.
- [16] M. Yuen and H. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal processing*, vol. 70, no. 3, pp. 247–278, 1998.
- [17] Z. Zhou, B. Wang, and J. Ma, "Scale-aware edge-preserving image filtering via iterative global optimization," *IEEE Transactions on Multimedia*, 2017.
- [18] Z. Wang, A. C. Bovik, and B. Evan, "Blind measurement of blocking artifacts in images," in *International Conference on Image Processing, 2000.*, vol. 3. IEEE, 2000, pp. 981–984.
- [19] X. Feng and J. P. Allebach, "Measurement of ringing artifacts in jpeg images," in *Proceedings of SPIE*, vol. 6076, 2006, pp. 74–83.
- [20] E. Bosc, P. L. Callet, L. Morin, and M. Pressigout, "An edge-based structural distortion indicator for the quality assessment of 3d synthesized views," in *2012 Picture Coding Symposium*, May 2012, pp. 249–252.
- [21] P. Hanhart and T. Ebrahimi, "Quality assessment of a stereo pair formed from decoded and synthesized views using objective metrics," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*. IEEE, 2012, pp. 1–4.
- [22] IVC-IRCCyN lab, "IRCCyN/IVC DIBR image database," [http://ivc.univ-nantes.fr/en/databases/DIBR\\_Images/](http://ivc.univ-nantes.fr/en/databases/DIBR_Images/), last accessed Aug. 30th 2017, [Online].
- [23] E. Bosc, R. Pepion, P. Le Callet, M. Koppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, "Towards a new quality metric for 3-d synthesized view assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1332–1343, 2011.
- [24] E. Bosc, P. Le Callet, L. Morin, and M. Pressigout, "Visual quality assessment of synthesized views in the context of 3d-tv," in *3D-TV system with depth-image-based rendering*. Springer, 2013, pp. 439–473.
- [25] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of graphics tools*, vol. 9, no. 1, pp. 23–34, 2004.
- [26] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3d warping using depth information for fvtv," *Signal Processing: Image Communication*, vol. 24, no. 1, pp. 65–72, 2009.
- [27] K. Mueller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View synthesis for advanced 3d video systems," *EURASIP Journal on Image and Video Processing*, vol. 2008, no. 1, pp. 1–11, 2009.
- [28] P. Ndjiki-Nya, M. Köppel, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand, "Depth image based rendering with advanced texture synthesis," in *2010 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2010, pp. 424–429.
- [29] P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3-d video," *IEEE Transactions on Multimedia*, vol. 13, no. 3, pp. 453–465, 2011.
- [30] M. Köppel, P. Ndjiki-Nya, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand, "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," in *2010 IEEE International Conference on Image Processing*. IEEE, 2010, pp. 1809–1812.
- [31] R. Song, H. Ko, and C. Kuo, "Mcl-3d: A database for stereoscopic image quality assessment using 2d-image-plus-depth source," *arXiv preprint arXiv:1405.1403*, 2014.
- [32] M. Solh and G. AlRegib, "Hierarchical hole-filling for depth-based view synthesis in ftv and 3d video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 495–504, 2012.
- [33] X. Liu, Y. Zhang, S. Hu, S. Kwong, C.-C. J. Kuo, and Q. Peng, "Subjective and objective video quality assessment of 3d synthesized views with texture/depth compression distortion," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4847–4861, 2015.

- [34] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of high efficiency video coding (hevc)," *IEEE Journal of selected topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, 2013.
- [35] Y. J. Jung, H. G. Kim, and Y. M. Ro, "Critical binocular asymmetry measure for the perceptual quality assessment of synthesized stereo 3d images in view synthesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 7, pp. 1201–1214, 2016.
- [36] D. Scharstein, R. Szeliski, and C. Pal, "Middlebury stereo datasets," 2012.
- [37] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on image processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
- [38] I. Ahn and C. Kim, "A novel depth-based virtual view synthesis method for free viewpoint video," *IEEE Transactions on Broadcasting*, vol. 59, no. 4, pp. 614–626, 2013.
- [39] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, "Reference softwares for depth estimation and view synthesis," *ISO/IEC JTC1/SC29/WG11 MPEG*, vol. 20081, p. M15377, 2008.
- [40] S. S. Yoon, H. Sohn, Y. J. Jung, and Y. M. Ro, "Inter-view consistent hole filling in view extrapolation for multi-view image generation," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2883–2887.
- [41] V. Jantet, C. Guillemot, and L. Morin, "Object-based layered depth images for improved virtual view synthesis in rate-constrained context," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 125–128.
- [42] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. 1–1.
- [43] O. Le Meur, J. Gautier, and C. Guillemot, "Exemplar-based inpainting based on local geometry," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 3401–3404.
- [44] G. Luo, Y. Zhu, Z. Li, and L. Zhang, "A hole filling approach based on background reconstruction for view synthesis in 3d video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1781–1789.
- [45] C. Zhu and S. Li, "Depth image based view synthesis: New insights and perspectives on hole generation and filling," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 82–93, 2016.
- [46] J.-L. Blin, "New quality evaluation method suited to multimedia context: Samviq," in *Proceedings of the Second International Workshop on Video Processing and Quality Metrics, VPQM*, vol. 6, 2006.
- [47] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP journal on image and video processing*, vol. 2008, no. 1, p. 659024, 2009.
- [48] I. ITU, "Methodology for the subjective assessment of video quality in multimedia applications," *Rapport technique, International Telecommunication Union*, 2007.
- [49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.
- [50] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2004.*, vol. 2. IEEE, 2003, pp. 1398–1402.
- [51] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [52] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, March 2002.
- [53] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on hvs," in *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, vol. 4, 2006.
- [54] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of dct basis functions," in *Proceedings of the third international workshop on video processing and quality metrics*, vol. 4, 2007.
- [55] D. Sandić-Stanković, D. Kukulj, and P. Le Callet, "Multi-scale synthesized view assessment based on morphological pyramids," *Journal of Electrical Engineering*, vol. 67, no. 1, pp. 3–11, 2016.
- [56] D. Sandić-Stanković, D. Kukulj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological wavelets," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2015, pp. 1–6.
- [57] D. Sandić-Stanković, D. Kukulj, and P. Le Callet, "DIBR-synthesized image quality assessment based on morphological multi-scale approach," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, p. 4, 2016.
- [58] P.-H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 82 881M–82 881M.
- [59] S. Tian, L. Zhang, L. Morin, and O. Deforges, "A full-reference image quality assessment metric for 3d synthesized views," in *Image Quality and System Performance Conference, at IS&T Electronic Imaging 2018*. Society for Imaging Science and Technology, 2018.
- [60] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, "Quality assessment of dibr-synthesized images by measuring local geometric distortions and global sharpness," *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 914–926, 2018.
- [61] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal processing letters*, vol. 17, no. 5, pp. 513–516, 2010.
- [62] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.
- [63] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.
- [64] K. Gu, V. Jakhetiya, J.-F. Qiao, X. Li, W. Lin, and D. Thalmann, "Model-based referenceless quality metric of 3d synthesized images using local image description," *IEEE Transactions on Image Processing*, 2017.
- [65] S. Tian, L. Zhang, L. Morin, and O. Déforges, "NIQSV: A no reference image quality assessment metric for 3D synthesized views," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, March 2017.
- [66] S. Tian, L. Zhang, L. Morin, and O. Déforges, "NIQSV+: A No-Reference Synthesized View Quality Assessment Metric," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1652–1664, 2018.
- [67] V. Q. E. Group, "Final report from the video quality experts group on the validation of objective models of multimedia quality assessment," *VQEG*, March 2008.