



HAL
open science

Perceived quality of DIBR-based synthesized views

Emilie Bosc, Romuald Pépion, Patrick Le Callet, Martin Köppel, Patrick Ndjiki-Nya, Luce Morin, Muriel Pressigout

► **To cite this version:**

Emilie Bosc, Romuald Pépion, Patrick Le Callet, Martin Köppel, Patrick Ndjiki-Nya, et al.. Perceived quality of DIBR-based synthesized views. SPIE Optics + Photonics, Aug 2011, San Diego, United States. pp.8135-16. hal-00617534

HAL Id: hal-00617534

<https://hal.science/hal-00617534>

Submitted on 2 Sep 2011

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.

Perceived quality of DIBR-based synthesized views

Emilie Bosc^a, Romuald P epion^b, Patrick Le Callet^b, Martin K oppel^c, Patrick Ndjiki-Nya^c, Luce Morin^a, Muriel Pressigout^a

^aIETR UMR CNRS 6164, INSA de Rennes, 20, avenue des Buttes de Coesmes, 35708 Rennes Cedex 7, France

^bIRCCyN, Universit  de Nantes, Rue Christian Pauc, 44306 Nantes, France

^cFraunhofer-Institut f r Nachrichtentechnik Heinrich-Hertz-Institut, Einsteinufer 37, 10587 Berlin, Germany

ABSTRACT

This paper considers the reliability of usual assessment methods when evaluating virtual synthesized views in the multi-view video context. Virtual views are generated from Depth Image Based Rendering (DIBR) algorithms. Because DIBR algorithms involve geometric transformations, new types of artifacts come up. The question regards the ability of commonly used methods to deal with such artifacts. This paper investigates how correlated usual metrics are to human judgment. The experiments consist in assessing seven different view synthesis algorithms by subjective and objective methods. Three different 3D video sequences are used in the tests. Resulting virtual synthesized sequences are assessed through objective metrics and subjective protocols. Results show that usual objective metrics can fail assessing synthesized views, in the sense of human judgment.

Keywords: Virtual view synthesis, multi-view video, 3DTV, quality assessment, quality metrics

1. INTRODUCTION

Any new device or method needs to be assessed through adapted and reliable tools. 3D Video technology has brought out new challenges such as the question of DIBR-based views evaluation. 3D Video applications, such as 3D Television (3DTV) or Free viewpoint Television (FTV) provide the users a novel type of experience thanks to the added value (depth, or immersion), compared to conventional 2D video. However, the success of these applications depends on their ability to offer high-quality visual content. FTV and 3DTV require virtual view generation, which is ensured by the use of DIBR algorithms. This process induces new types of artifacts and consequently impacts on quality that has to be identified considering various contexts of use. Artifacts from 2D video compression have been widely studied, but DIBR deals with new types of distortions that are mainly geometric distortions.

Yet, synthesized views, generated either from decoded and distorted data or from original data, need to be assessed. The best assessment tool remains the human judgment as long as the right protocol is used. Subjective quality assessment is still delicate while addressing new types of conditions because one has to define the optimal way to get reliable data. Subjective tests are time-consuming consequently one should draw big lines on how to conduct such experiment to save time and observers. Since DIBR is introducing new conditions, the right protocols to assess the visual quality with observers are still an open question. Protocols might vary according to the purpose (impact of compression, DIBR techniques comparison ...).

Objective metrics are requested because subjective quality assessment tests are cumbersome. They are meant to predict human judgment and to facilitate the quality evaluation. However, the evaluation of their performances relies on their closeness to subjective assessment results. As, the way to conduct the subjective quality assessment protocols is already questionable, reliable objective quality metrics among existing ones that could be useful in DIBR context are all the more uncertain.

Yet, trustworthy working groups partly base their future specifications, concerning new strategies for 3D video, on the outcome of objective metrics. Considering the test conditions may rely on usual subjective and objective protocols (because of their availability), the outcome of wrong choices could result to a poor quality of experience for users, or worse, the failure of 3D Video for industrials.

This study is motivated by two main questions: are usual requirements for subjective evaluation protocols still appropriate for assessing 3D synthesized views? And second, are usual metrics sufficient for assessing 3D synthesized views?

Previous studies (^{1,2}) already proposed to answer our second questioning, regarding the reliability of usual objective metrics. However, often, objective metrics are applied on single distorted views and these measurements are compared to subjective assessments on 3D displays whose synthesis process is blind. In ², artifacts from both synthesis and compression are assessed without distinction.

This study aims at answering the two questions above, but considering the complexity of 3D Video, and in order to precisely keep in check what is being measured, a first approach is to be independent from the 3D, that is to say both the stereopsis and the 3D display whose technology is still a major factor of quality degradation. Thus, the study presented in this chapter focuses on the quality evaluation of DIBR-based synthesized views in 2D conditions. Besides, these conditions are plausible in a Free Viewpoint Video (FVV) application.

The problem we consider has not been treated in such a way in the literature. In our context, the results of our study confirm the common idea that 3D video is not just the extension of 2D video and that commonly used 2D video metrics are not sufficient to assess the synthesized views quality. The synthesized images quality is measured through pixel-based or perceptual-like metrics. The images are synthesized with seven different DIBR algorithms. Correlation with human perception is estimated thanks to subjective assessments.

The document is organized as follows: Section 2 highlights the challenges brought by 3D in the context of quality assessment. Section 3 presents the material used for the experiments. Section 4 considers the first question regarding the subjective assessment protocols while Section 5 studies the reliability of the objective tools. Section 6 concludes the paper.

2. NEW CHALLENGES IN TERMS OF QUALITY ASSESSMENT

Recently, the complexity of 3D video has been considered. It is now admitted that 3D Video is not the trivial extension of 2D to 3D, thanks to studies such as Seuntiens³. In the latter, Seuntiens claims that not only the image quality should be addressed in a quality evaluation, but as well new factors such as naturalness, visual experience, and presence. This suggests the use of a multidimensional measurement that does not exist yet as an objective tool, and that has not been deployed yet in subjective quality assessment of 3D media.

Indeed, it is a tough task designing subjective quality assessment frameworks, when new concepts are introduced. Besides, Chen et al.⁴ proposed new requirements from this perspective. They proposed to reconsider several conditions in this context, such as the viewing conditions (viewing distance, monitor resolution), the test material (depth rendering according to the chosen 3D display), viewing duration, etc. This certainly reinforces our assumption concerning the fact that 3D Video is much more complex than 2D conventional video, and that tools usually thought of as appropriate in the context of 2D conventional video should be reconsidered.

As a matter of fact, the study of artifacts occurring when generating a novel view from DIBR algorithms reveals that the distortions are mainly located around strong discontinuities, transitions of depth. They are quite different from artifacts we are used to deal with in 2D video compression. This comes from the fact they arise from the very synthesis process, and are then geometric distortions.

These distortions are directly related to the quality of the incoming data values (e.g. quality of depth estimation) and the synthesis process strategies. The latter mainly aimed at dealing with the critical problem in DIBR, namely the disocclusion. This term refers to the discovered regions during the synthesis of a new viewpoint. Extrapolation techniques are meant to fill the disoccluded regions. In this section, we comment the artifacts from seven different synthesis process strategies. These seven DIBR algorithms are referenced from A1 to A7:

- A1: based on Fehn⁵, where the depth map is pre-processed by a low-pass filter. Borders are cropped, and then an interpolation is processed to reach the original size.
- A2: based on Fehn⁵. Borders are inpainted by the method proposed by Telea⁶.

- A3: Tanimoto et al. ⁷, it is the recently adopted reference software for the experiments in the 3D Video group of MPEG.
- A4: Müller et al. ⁸, proposed a hole filling method aided by depth information.
- A5: Ndjiki-Nya et al. ⁹, the hole filling method is a patch-based texture synthesis.
- A6: Köppel et al. ¹⁰, uses depth temporal information to improve the synthesis in the disoccluded areas.
- A7: corresponds to the unfilled sequences (i.e. with holes).

Commonly encountered artifacts, in DIBR-based synthesized images are described in the following. In most of the cases, these artifacts are located around strong discontinuities, but are more perceptible when the contrast between the background and the foreground is important.

Object shifting: a region may be slightly translated or resized, depending on the chosen extrapolation method (if the method chooses to assign the background values to the missing areas, object may be resized), or on the encoding method (blocking artifacts in depth data results in object shifting in syntheses).

Blurry edges: when data is pre-processed by low pass filters, it leads to blurry synthesized regions. Indeed, to avoid, the lack of information in disoccluded areas, it is tempting to low pass filter the per-pixel depth information, as proposed in A1. However, this smoothes out the discontinuities and results in blurry regions in synthesis. This is annoying especially around the edges that are expected to be sharp. It is obvious around the background/foreground transitions. These remarks are confirmed on Figure 1-A1, where we also note the curved stairway, and on Figure 1-A3 around the disoccluded areas.

Texture synthesis: inpainting methods can fail in filling complex textured areas. A2 and A3 use the inpainting method in ⁶. However, this inpainting method fails to reconstruct complex textures. A4 also fails in rendering complex textures. To overcome these limitations, a hole filling approach based on patch-based texture synthesis is proposed in A5. However, candidate patches may not be perceptually close to the disoccluded region and then lead to artifacts. Furthermore, the rectangular patches can lead to blocky artifacts.

Flickering: when errors occur randomly in depth data along the sequence, pixels are wrongly projected: some pixels suffer slight changes of depth, which appears as flickers in the resulting synthesized pixels. To avoid this effect, proposed the use of a background sprite is proposed in A6. Background image information is stored in the sprite and updating by the original and the synthesized image information of previous frames to provide depth consistency. All along the frames, background texture is copied from the stored information to the disoccluded region. The copy relies on depth data, and is then still dependent on depth data quality

Tiny distortions: in synthesized sequences, a large number of tiny geometric distortions and illumination differences are temporally constant and perceptually invisible. However, pixel-based metrics may penalize these distorted zones.

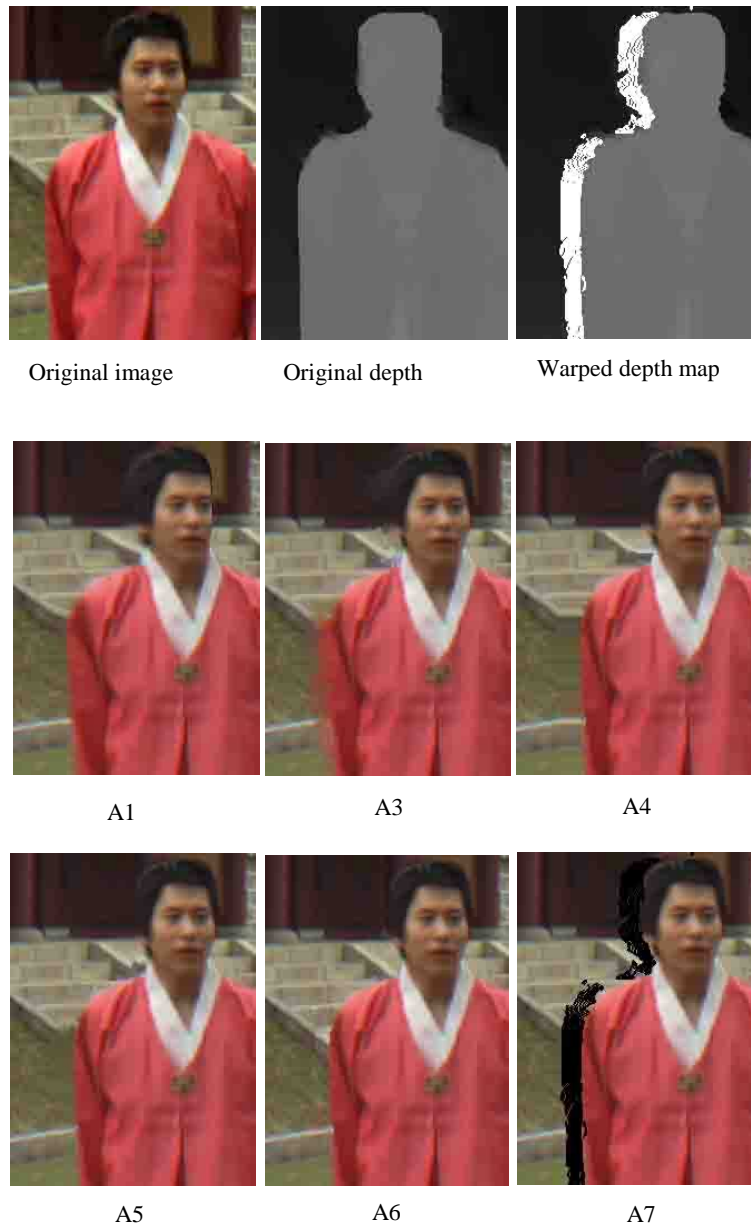


Figure 1: DIBR-based synthesized frames of the "Lovebird1" sequence.

3. EXPERIMENTAL MATERIAL

3D Video involves many factors that are not completely under control yet. Among them, the stereopsis phenomenon, or the relationship between depth rendering and the quality of experience over 3D-adapted displays, can be cited. To be independent of combination of uncontrolled parameters, the quality of images synthesized from DIBR algorithms is first assessed in 2D conventional conditions. This is meant to be a preliminary experiment, and it is also a plausible context for an FVV application.

Three different MVD sequences are used. The sequences are *Book Arrival* (1024x768, 16 cameras with 6.5cm spacing), *Lovebird1* (1024x768, 12 cameras with 3.5 cm spacing) and *Newspaper* (1024x768, 9 cameras with 5 cm spacing). These sequences are also involved in the test benchmarks of 3DV Group of MPEG.

Seven DIBR algorithms processed the three MVD sequences to generate, for each one, four different viewpoints. Six of the DIBR algorithms, from A1 to A6, are cited in section 2, and the seventh correspond to the unfilled sequences (i.e. with holes) is labelled A7.

Subjective and objective evaluations were processed to evaluate the synthesized views quality.

As for the subjective quality evaluations, standardized methodologies for subjective multimedia quality assessment, such as Absolute Categorical Rating (ACR)¹¹, have proved their efficiency regarding the quality evaluation of 2D conventional images. ACR was then used to collect perceived quality scores. The tests were conducted in an ITU conforming test environment. The stimuli were displayed on a TVLogic LVM401W, and according to ITU-T BT.500¹².

As for the objective quality metrics, measurements were realized over the 84 synthesized views by the means of MetriX MuX Visual Quality Assessment Package¹³ except for two metrics: VQM and VSSIM. VQM were available at¹⁴; VSSIM was implemented by the authors, according to¹⁵. The reference was the original acquired video. Still image metrics were applied on each frames of the sequences and then averaged by the number of frames.

4. SUITABILITY OF SUBJECTIVE QUALITY ASSESSMENT PROTOCOLS

ACR is selected as an example of subjective assessment protocol to reconsider. ACR methodology consists in presenting test objects (i.e. images or sequences), one at a time and in a random order, to the observers. Observers score the test item according to a discrete category rating scale. This quality scale is depicted in

Table 1: ACR quality scale.

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

A semantic meaning is assigned to the five integers of the quality scale. The results of an ACR test are obtained by averaging observers' opinion scores for each stimulus, in other words, by computing mean opinion scores (MOS). ACR requires many observers to minimize the contextual effects (previously presented stimuli influence the observer opinion, i.e. presentation order influences opinion ratings). The accuracy increases with the number of participants. The efficiency of ACR methodology is proved as soon as at least 24 observers participate, according to VQEG.

ACR test was conducted over the 84 video sequences synthesized from the seven DIBR algorithms.

Table 2 shows the ranking of the algorithms from the obtained subjective scores. The obtained values indicate the quality of the presented sequences is perceived as "poor", in average.

Table 2: Ranking of algorithms according to subjective scores.

	A1	A2	A3	A4	A5	A6	A7
ACR	2.70	2.41	2.14	2.03	1.96	2.13	1.28
Rank order	1	2	3	5	6	4	7

Although ACR scores allow the ranking of the algorithms, they do not directly provide knowledge on the statistical equivalence of the results. Student's t-test has been performed over the scores to meet this need.

Table 1 depicts the results of the Student's t-test processed with the values. The number in parentheses indicates the required minimum number of observers that allow the statistical distinction (VQEG recommends 24 participants as a minimum¹⁶, values in bold are higher than 24 in the table). It seems that in some cases, 24 participants are not sufficient to provide a sharp distinction. This may be due to the fact that some of the algorithms induce the same type of distortions. Then, the perceived quality is similar to observers.

Table 3: Results of Student's t-test with ACR results. Legend: ↑: superior, ↓: inferior, O: statistically equivalent. Reading: Line "1" is statistically superior to column "2". Distinction is stable when "less than 7 observers participate.

	A1	A2	A3	A4	A5	A6	A7
A1		↑(7)	↑(3)	↑(3)	↑(2)	↑(3)	↑(1)
A2	↓(7)		↑(2)	↑(2)	↑(1)	↑(2)	↑(1)
A3	↓(3)	↓(2)		^o (>32)	↑(9)	^o (>32)	↑(1)
A4	↓(3)	↓(2)	^o (>32)		^o (>32)	^o (>32)	↑(1)
A5	↓(2)	↓(1)	↓(9)	^o (>32)		↓(15)	↑(1)
A6	↓(3)	↓(2)	^o (>32)	^o (>32)	↑(15)		↑(1)
A7	↓(1)	↓(1)	↓(1)	↓(1)	↓(1)	↓(1)	

5. RELIABILITY OF OBJECTIVE METRICS

Different objective metrics are used in the presented experiments:

- Peak Signal to Noise Ratio (PSNR) measures the signal fidelity of a distorted image compared to a reference. It is based on the measure of the Mean Squared Error (MSE).
- As an alternative to pixel-based methods, Universal Quality Index UQI¹⁷ is a perceptual-like metric. It models the image distortion by a combination of three factors: loss of correlation, luminance distortion, and contrast distortion.
- PSNR-HVS¹⁸, based on PSNR and UQI, is meant to take into account the Human Visual System (HVS) properties.
- PSNR-HVSM¹⁹ is based on PSNR but takes into account Contrast Sensitivity Function (CSF) and between-coefficient contrast masking of DCT basis functions.
- Single-scale Structural SIMilarity (SSIM)²⁰ combines image structural information: mean, variance, covariance of pixels, for a single local patch. Multi-scale SSIM (MSSIM) is the average SSIM scores of all patches of the image.
- Video Structural Similarity Measure (V-SSIM)¹⁵ is a FR video quality metric which uses structural distortion as an estimate of perceived visual distortion. At the patch level, SSIM score is a weighted function of SSIM of the different component of the image (i.e. luminance, and chromas). At the frame level, SSIM score is a weighted function of patches' SSIM scores (based on the darkness of the patch). Finally at the sequence level, VSSIM score is a weighted function of frames' SSIM scores (based on the motion).
- Visual Signal to Noise Ratio (VSNR)²¹ is also a perceptual-like metric: it is based on a visual detection of distortion criterion, helped by CSF.
- Weighted Signal to Noise Ratio (WSNR) that uses a weighting function adapted to HVS denotes a weighted Signal to Noise Ratio as applied in²² and expressed in²³.
- Visual Information Fidelity (VIF)²⁴ uses a statistical model and an image distortion model to assess an image. VIFP is a pixel-based version of VIF.
- Information Fidelity Criterion (IFC)²⁵ uses a distortion model to evaluate the information shared between the reference image and the degraded image.

- Noise quality measure (NQM) ²² quantifies the injected noise in the tested image.
- Video Quality Metric (VQM) was proposed by Pinson and Wolf in ²⁶. VQM combines different perceptual calculated parameters (motion, blurring, global noise, color distortion) to compute the final quality measure.

Table 4 shows the rankings obtained from the objective scores (and the subjective scores as a reminder). Subjective-based ranking points A1 as the best out of the seven algorithms. But objective-based rankings judge it as the worst. As well, A6 obtains the best results for numerous objective metrics, but subjective assessments judge the synthesized views as “poor”. From the observation of this table, our assumption is that objective metrics detect non-annoying artifacts. Note, that objective metrics are coherent to one another. They are correlated.

The consistency between objective and subjective measures is evaluated by computing the correlation coefficients for the whole fitted measured points.

Table 5 shows the results. None of the tested metric is 50% close to human judgment when assessing synthesized views. According to MOS and PC scores, the most correlated metrics are VSNR, VIFP and IFC (47.3%, 46.9%, and 45.6%).

In order to consider further the consistency, Figure 2 shows the ACR values over the VSNR scores. The figure shows that a difference of nearly 3dB (A3 and A6) leads to nearly the same ACR score. However the range of confidence intervals is quite important for cases such as A3’s. This can be explained by the fact that the results presented in Figure 2 are obtained from the twelve synthesized sequences of each algorithm whose quality scores varied depending on the baseline distance between reference cameras and target viewpoints, and also depending on the sequences.

Table 4: Rankings according to measurements.

	A1	A2	A3	A4	A5	A6	A7
ACR	2.70	2.41	2.14	2.03	1.96	2.13	1.28
Rank order	1	2	3	5	6	4	7
PSNR	19.02	24.99	23.227	25.994	26.035	26.04	20.89
Rank order	7	4	5	3	2	1	6
SSIM	0.648	0.844	0.786	0.859	0.859	0.859	0.824
Rank order	7	4	6	1	1	1	5
MSSIM	0.664	0.932	0.825	0.948	0.948	0.948	0.888
Rank order	7	4	6	1	1	1	5
VSNR	13.14	20.41	18.75	21.786	21.965	21.968	20.73
Rank order	7	5	6	3	2	1	4
VIF	0.129	0.393	0.313	0.423	0.423	0.424	0.396
Rank order	7	5	6	2	2	1	4
VIFP	0.153	0.415	0.342	0.446	0.446	0.446	0.419
Rank order	7	5	6	1	1	1	4
UQI	0.359	0.664	0.58	0.598	0.598	0.598	0.667
Rank order	7	5	6	3	3	3	1
IFC	0.779	2.399	1.926	2.562	2.562	2.564	2.404
Rank order	7	5	6	2	2	1	4
NQM	8.66	15.933	13.415	16.635	16.739	16.739	10.63
Rank order	7	4	5	3	1	1	6
WSNR	14.41	20.85	18.853	21.76	21.839	21.844	16.46
Rank order	7	4	5	3	2	1	6
PSNR HSV	13.99	19.37	18.361	21.278	21.318	21.326	16.23
Rank order	7	4	5	3	2	1	6
PSNR HSV	13.74	19.52	17.958	20.795	20.823	20.833	15.91
Rank order	7	4	5	3	2	1	6
VSSIM	0.662	0.879	0.809	0.899	0.898	0.893	0.854
Rank order	7	4	6	1	2	3	5
VQM	0.888	0.623	0.581	0.572	0.556	0.557	0.652
Rank order	7	5	4	3	1	2	6

Table 5: Correlation coefficient between objective and subjective scores.

	PSNR	SSIM	MSSIM	VSNR	VIF	VIFP	UQI	IFC	NQM	WSNR	PSNR HSV	PSNR HVS	VSSI M	VQM
ACR	34.5	45.2	27	47.3	43.9	46.9	20.2	45.6	36.6	32.9	34.5	33.9	33	33.6

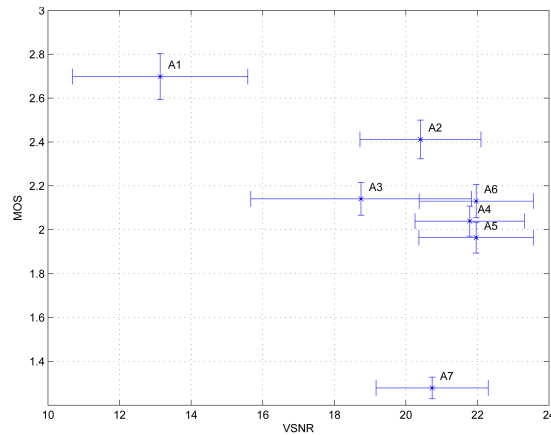


Figure 2: Correlation between MOS and VSNR.

These results show inconsistencies between subjective and objective scores. This suggests that objective metrics are not able to reliably predict human judgment, in the tested conditions.

6. CONCLUSION

This paper addresses the issue of evaluating virtual synthesized views with tools usually appropriate in the context of 2D conventional video quality assessment. The assessments of the seven test algorithms by objective measurements and subjective ratings show that among all tested objective metrics, VSNR VIFP and IFC are the most correlated with perceptual evaluation provided by MOS scores, with low correlation scores though. As well, the results also show metric's inability to predict human experience, in the tested conditions. The objective metrics certainly indicated the presence of distortions, but they were not able to predict annoying distortions in the sense of human opinion. New methods are then required for assessing virtual synthesized views as pixel-based and perceptual-based metrics fail. Because view synthesis induces geometric distortions, a new model should take into account depth data and particularly the preservation of coherent edges after the synthesis.

7. ACKNOWLEDGEMENTS

We would like to acknowledge the Interactive Visual Media Group of Microsoft Research for providing the Breakdancers data set, the MPEG Korea Forum for providing the Lovebird1 data set, the GIST for providing the Newspaper data set, and HHI for providing Book Arrival.

This work is supported by the French National Research Agency as part of PERSEE project (ANR-09-BLAN-0170) and CAIMAN project (ANR-08-VERS-002).

REFERENCES

- [1] Yasakethu, S.L.P., Hewage, C., Fernando, W., and Konoz, A., "Quality analysis for 3D video using 2D video quality models," *Consumer Electronics, IEEE Transactions on* 54(4), 1969-1976 (2008).
- [2] Tikanmaki, A., Gotchev, A., Smolic, A., and Mueller, K., "Quality assessment of 3D video in rate allocation experiments," in *IEEE Int. Symposium on Consumer Electronics (14-16 April, Algarve, Portugal)* (2008).
- [3] Seuntiens, P., "Visual Experience of 3D TV," *Doctoral thesis, Eindhoven University of Technology* (2006).
- [4] Chen, W., Fournier, J., Barkowsky, M., and Le Callet, P., "New requirements of subjective video quality assessment methodologies for 3DTV" (2010).
- [5] Fehn, C., and others, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI* 5291, 93-104 (2004).
- [6] Telea, A., "An Image Inpainting Technique Based on the Fast Marching Method," *Journal of Graphics, GPU, and Game Tools* 9(1), 23-34 (2004).

- [7] Tanimoto, M., Fujii, T., Suzuki, K., Fukushima, N., and Mori, Y., "Reference Softwares for Depth Estimation and View Synthesis," presented at ISO/IEC JTC1/SC29/WG11 MPEG 2008/M15377, April 2008.
- [8] Müller, K., Smolic, A., Dix, K., Merkle, P., Kauff, P., and Wiegand, T., "View synthesis for advanced 3D video systems," *EURASIP Journal on Image and Video Processing* (2008).
- [9] Ndjiki-Nya, P., Köppel, M., Doshkov, D., Lakshman, H., Merkle, P., Müller, K., and Wiegand, T., "Depth image based rendering with advanced texture synthesis," in *Proc. IEEE International Conference on Multimedia & Expo (ICME)* (2010).
- [10] Köppel, M., Ndjiki-Nya, P., Doshkov, D., Lakshman, H., Merkle, P., Müller, K., and Wiegand, T., "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," in *Proc. IEEE International Conference on Image Processing (ICIP)* (2010).
- [11] ITU-T Study Group 12, "ITU-T P.910 Subjective video quality assessment methods for multimedia applications" (1997).
- [12] ITU-R BT., [500, Methodology for the subjective assessment of the quality of television pictures] , November (1993).
- [13] "MetriX MuX Home Page," <http://foulard.ece.cornell.edu/gaubatz/metrix_mux/>.
- [14] "Video Quality Research," <<http://www.its.bldrdoc.gov/vqm/>>.
- [15] Wang, Z., Lu, L., and Bovik, A., "Video quality assessment based on structural distortion measurement," *Signal processing: Image communication* 19(2), 121-132 (2004).
- [16] VQEG 3DTV Group, "VQEG 3DTV Test Plan for Crosstalk Influences on user Quality of Experience" (2010).
- [17] Wang, Z., and Bovik, A.C., "A universal image quality index," *Signal Processing Letters, IEEE* 9(3), 81-84 (2002).
- [18] Egiazarian, K., Astola, J., Ponomarenko, N., Lukin, V., Battisti, F., and Carli, M., "New full-reference quality metrics based on HVS," in *CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics* (2006).
- [19] Ponomarenko, N., Silvestri, F., Egiazarian, K., Carli, M., Astola, J., and Lukin, V., "On between-coefficient contrast masking of DCT basis functions," in *CD-ROM Proc. of the Third International Workshop on Video Processing and Quality Metrics 4* (2007).
- [20] Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P., "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on* 13(4), 600-612 (2004).
- [21] Chandler, D.M., and Hemami, S.S., "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *Image Processing, IEEE Transactions on* 16(9), 2284-2298 (2007).
- [22] Damera-Venkata, N., Kite, T.D., Geisler, W.S., Evans, B.L., and Bovik, A.C., "Image quality assessment based on a degradation model," *Image Processing, IEEE Transactions on* 9(4), 636-650 (2002).
- [23] Lee, S., Pattichis, M.S., and Bovik, A.C., "Foveated video quality assessment," *Multimedia, IEEE Transactions on* 4(1), 129-132 (2002).
- [24] Sheikh, H.R., and Bovik, A.C., "Image information and visual quality," *Image Processing, IEEE Transactions on* 15(2), 430-444 (2006).
- [25] Sheikh, H.R., Bovik, A.C., and de Veciana, G., "An information fidelity criterion for image quality assessment using natural scene statistics," *Image Processing, IEEE Transactions on* 14(12), 2117-2128 (2005).
- [26] Webster, A.A., Jones, C.T., Pinson, M.H., Voran, S.D., and Wolf, S., "An objective video quality assessment system based on human perception," in *SPIE Human Vision, Visual Processing, and Digital Display IV* 1913, 15-26 (1993).