Using region-of-interest for quality evaluation of DIBR-based view synthesis methods

Andrei Purica*†, Giuseppe Valenzise*, Beatrice Pesquet-Popescu* and Frederic Dufaux*
*LTCI, CNRS, Telecom ParisTech, Universite Paris-Saclay, 75013, Paris, France
†University Politehnica of Bucharest, 061071, Romania
Email: {purica, valenzise, pesquet, dufaux}@telecom-paristech.fr

***Abstract*—As 3D media became more and more popular over the last years, new technologies are needed in the transmission, compression and creation of 3D content. One of the most commonly used techniques for aiding with the compression and creation of 3D content is known as view synthesis. The most effective class of view synthesis algorithms are using Depth-Image-Based-Rendering techniques, which use explicit scene geometry to render new views. However, these methods may produce geometrical distortions and localized artifacts which are difficult to evaluate as they are inherently different from encoding errors and they are perceived differently by human subjects. In this paper, we propose a region-of-interest evaluation technique for view synthesis based on DIBR methods. Based on the assumption that certain areas determined by the geometrical properties of the scene are prone to distortions, we select a ROI by analyzing the multiple DIBR methods together with the ground truth. The approach is tested using a subjective evaluation view synthesis database and show that our method improves the SSIM correlation with subjective scores We also test another similar method and traditional metrics.**

Keywords*—*View Synthesis; Visual quality assessment; Multi-View Video; SSIM; Depth-Image-Based-Rendering

I. Introduction

In the past years 3D content has become more and more popular. As display technology advances, 3D content is now available to the large public and can be enjoyed on most modern television sets. This created a need for efficient compression and transmission systems for 3D information and new algorithms for the creation of 3D movies or 2D to 3D conversion. Some of the most common applications that involve 3D information are free view point television (FTV) [1], immersive teleconference systems, medical applications and gaming [2].

Several new formats for 3D information representation exist. Some of the most popular include stereo video, MultiView Video (MVV) and Multiview-Video-plus-Depth (MVD) [3]. While stereo video allows a user to experience the sensation of depth, MVV and MVD formats provide additional options such as changing the point of view on the scene or varying the perception of depth. The latter two formats both consist of a number of video sequences acquired in parallel at different points of view of the same scene. In the case of MVD, these texture sequences are also accompanied by depth information in the form of depth maps. Because of this, MVD format supports the creation of virtual view points of the scene by means of Depth-Image-Based-Rendering (DIBR) techniques [4].

The process of generating a video sequence or an image as if acquired from a new point of view from existing sequences or images is known as *view synthesis*. Several methods exist in the literature and can be mainly divided into three categories based on the use of geometrical information: *i*) Methods that do not require geometrical information and use interpolation and filtering to synthesize new views. Some of the most popular ones include light field rendering [5], concentric mosaics [6] or lumigraph [7], *ii*) Methods that use implicit geometry such as pixel correspondences computed with optical flow or any other motion estimation technique [8], *iii*) and finally, methods that use explicit scene geometry in the form of depth maps, to warp pixels from one view into a virtual one [9] [10]. The later category received great interest as it provides a fast and efficient way of generating multiple views.

In the past years, the Moving Picture Experts Group (MPEG) began developing a 3D extension of the High Efficiency Video Coding (HEVC) standard [11], to meet the need for an MVD coding standard. An experimental framework for 3D video was developed [12] and a 3D-HEVC test model (3D-HTM) [13] was build. The model also incorporates a View Synthesis Reference Software (VSRS-1DFast), which uses DIBR techniques to render new views from the texture information and associated depth maps.

However, the quality of the virtual views is greatly affected by multiple factors. A first issue is related to areas in the virtual view which are not visible in the reference views, as no information is available and they manifest as holes in the synthesized image. These areas are also known as disocclusions. They can be divided in two types based on their location: border or non-border disocclusions [14]. The first category are produced by the displacement of the field of view and are located on the sides of the images. The second category appears around foreground object edges. In order to avoid non-border disocclusions it is usually preferred to merge two synthesized views from a left and a right reference view. However, parts of the non-border disocclusions may coincide in the merged views. Traditionally this problem is resolved using inpainting algorithms such as [15] [16] [17]. Other methods propose a preprocessing of the depth maps in order to reduce the size of disocclusions [18], [19]. When working on video sequences, temporal correlations can also be exploited to retrieve information on disoccluded areas [20] or a background extraction can be performed [21] [22].

Other types of artifacts specific to DIBR methods are caused by the depth maps quality. In addition to coding artifacts, depth maps are generally not perfect and may contain noise. This can lead to different types of artifacts in the synthesized image. A common problem is the texture-depth alignment which may lead to pixels belonging to a foreground object to be warped as if they are part of the background or vice-versa. Another issue is related to the precision of the depth which may cause small ”cracks” in the synthesized image due to incorrect displacement of pixels. Finally, depths maps are also subjected to a quantization process as real depth values are not actually stored, they are usually quantized to 256 levels. In general, this problems appear in areas where depth maps are not uniform (i.e. foreground/background separation).

Because the artifacts produced by synthesis are inherently different from those of encoding, evaluating the quality of synthesis in systems using DIBR rendering is not a trivial matter. Especially, considering the final goal of such systems is to provide a 3D experience. The Video Quality Expert Group (VQEG) created the 3DTV Work Group, which is now part of the Immersive Media Group [23], to conduct experiments on the quality of 3D media. Numerous studies were made to address the problem of synthesized video evaluation. Tikanmaki *et al.* [24] studied the assessment of 3-D encoded video, the authors also considered the synthesized view quality. Bosc *et al.* [25] studied the quality of DIBR synthesis and proposed two approaches based on a region of interest (ROI) evaluation. A first method analyzes the contours shifts in the synthesized view and a second one focuses on evaluating the mean SSIM score over disoccluded areas. Purica *et al.* [26] study the difference between encoding and synthesis artifacts and propose a ROI based SSIM by separating between encoding errors coming from the reference view and distortion caused by the DIBR warping process. In this paper we extend the ideas presented in [26] and propose a new ROI generation technique for SSIM evaluation of synthesized videos. Next, we perform a study of the results using a subjective evaluation database in order to validate our assumptions.

The rest of this paper is organized as follows. Section II motivates and shows the proposed evaluation technique. Section III-A describes the subjective evaluation database used for validating this technique. In Sections III-B and III-C we describe our testing methodology and report our experimental results. Finally, Section IV concludes the paper.

II. Towards a region of interest evaluation

As discussed in Sec. I synthesized videos can have multiple types of artifacts which affect the quality of the image in different ways. DIBR synthesis methods compute pixel disparity, from depth map sequences, and then warp the images from the reference view into a new view. Depth maps are usually stored as video sequences and the values are inversely quantized to 256 levels with respect to real scene depth. In the case of aligned camera systems the disparity is easily obtained using the following equation:

$$\mathbf{d}(\mathbf{k}) = f \cdot B \left[\frac{Z(\mathbf{k})}{255} \left(\frac{1}{Z_{min}} - \frac{1}{Z_{max}} \right) + \frac{1}{Z_{max}} \right] \quad (1)$$

where f is the focal length of the camera, B is the distance between view points, Z_{min}, Z_{max} are the minimum and maximum depth of the scene and $\mathbf{k} = (x, y)$ is a position in the image. Because depth maps are subjected to distortions from the acquiring device or transmission systems, the synthesized image can be subjected to geometrical distortion of foreground objects and also poor reproduction of complex textures. As noted in other studies [26] [25] [27], traditional metrics such as PSNR or SSIM may not be the best way to asses the quality of synthesized images. This behavior can be explained by the strong correlation between scene geometry and position of highly distorted areas. In Figure 1 we depict

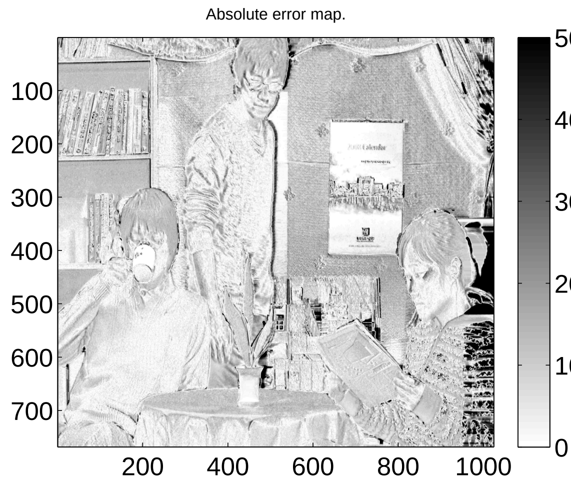


Fig. 1. Absolute error color map for frame 93 of Newspaper sequence. View 6 synthesized from view 4 using [4].

a gray scale representation of the absolute errors of frame 93 of Newspaper sequence synthesized using [4]. Black indicates an absolute error higher than 50 while white represents an absolute error of 0. It is easily noticeable that the absolute errors are not uniformly distributed throughout the image and are concentrated in certain critical areas. In this example view 6 was synthesized from view 4. We can see a large concentration of high errors on the left side of the image. This is consistent to a border disocclusion which was filled with an inpainting algorithm. Furthermore, highest errors are concentrated around foreground objects and there exists a high correlation between scene geometry and high distortions. Areas that have the same depth and uniform textures are usually represented without distortions, while foreground object edges and more complex textures have a high distortion. Also, we can notice that not all contours are equally distorted. In this example right most edges of objects tend to have a higher distortion. This behavior can be attributed to the direction of the synthesis from view 4 to view 6, which results in holes on one side of the foreground objects. This type of spatial error distribution is usually similar in most DIBR methods. Because of this, using a ROI when evaluating the quality of synthesis methods may provide a better indication of a method’s performance.

Given the goal of evaluating multiple synthesis methods the ROI can be selected as discussed in Section I by thresholding the absolute error or analyzing contours. Another possibility which may provide good results is to look at areas that are rendered differently by the methods which we want to evaluate. This is a reasonable assumption as background areas with non complex texture are usually identical in most synthesis methods and do not affect the quality of the image. Also areas that are rendered identically by multiple methods do not provide any differentiation between the tested DIBR algorithms.

Consider a number of distorted images $\mathbf{I_1^d}, \mathbf{I_2^d}, .., \mathbf{I_n^d}$. Each image is a synthesis of the same view using the same reference and one of n methods. We define $\mathcal{P}$ as:

$$\mathcal{P}(x,y) = \text{std}([\mathbf{I_1^d}(x,y), \mathbf{I_2^d}(x,y), .., \mathbf{I_n^d}(x,y)]) \qquad (2)$$

where (x, y) denotes a position in the image and std is the standard deviation.

The binary mask of the ROI can be expressed as:

$$B(x,y) = \begin{cases} 1 & \text{if } \mathcal{P}(x,y) > \tau \cdot \text{mean}(\mathcal{P}) \\ 0 & \text{if otherwise} \end{cases} \qquad (3)$$

where τ is a coefficient used to balance the selection and mean is the average value of $\mathcal{P}$.

As the ground truth is also available when computing the ROI, it is possible to include it in the computation. Including the ground truth in the computation does not provide information towards differentiating the methods. However, it may lead to a more balanced selection of critical areas by taking into account not only regions which differ in the tested methods but also regions that have a relatively high distortion in all methods. This way, the score will also reflect the global quality of a synthesized image instead of only with respect to the tested methods.

$$\mathcal{P}(x,y) = \text{std}([\mathbf{I_1^d}(x,y), .., \mathbf{I_n^d}(x,y), \mathbf{I_1^r}(x,y), .., \mathbf{I_m^r}(x,y)]) \qquad (4)$$

where I^r is the reference used to compute the metric and m is the number of times we add the ground truth. Due to a variable number of methods that can be evaluated in parallel, the ground truth needs to be weighted. In our experiments we used a weight of $1/6$ (i.e. the ground truth was added once). However, in this case, the mask will have a lot of noise in the form of localized pixels selected for evaluation. Because the artifacts depend on the structure of the scene it is best to remove single pixels and also consider the neighborhood of the critical areas. This can be achieved by performing an erosion and dilation operation on the binary mask. In order to extend the initial ROI, the dilation operation should use a larger morphological structuring element. In our tests we used a $2x2$ square element for the erosion and a $7x7$ square element for the dilation. This values were selected empirically.

III. Experimental Results

In this section we report our findings using the ROI evaluation technique described in Section II and use a subjective evaluation test database to validate the results. The first part of this section will describe the subjective evaluation database. The second part describes the methodology and finally, the results are discussed in the last part.

TABLE I. Sequences used in our experiments

Sequence	**Resolution**	**Frames per second**	**Number of frames**	Views
Book arrival	1024×768	15	100	8 9 10
Lovebird	1024×768	30	150	6 7 8
Newspaper	1024×768	30	200	4 5 6

A. Subjective evaluation database used in our experiments

In order to validate this technique we use a view synthesis subjective evaluation database available at [28]. The tests were performed using Absolute Categorical Rating with Hidden Reference Removal (ACR-HR) [29] with 32 subjects. Three multiview video sequences were used: Book arrival, Lovebird, Newspaper. Sequence details are reported in Table I. For each sequence there are three views used in the experiments: a left, center and right view indicated in Table I. Four synthesized views are generated for each sequence: left→right, right→left, left→center, right→center. Each synthesis is then performed using the seven methods described below:

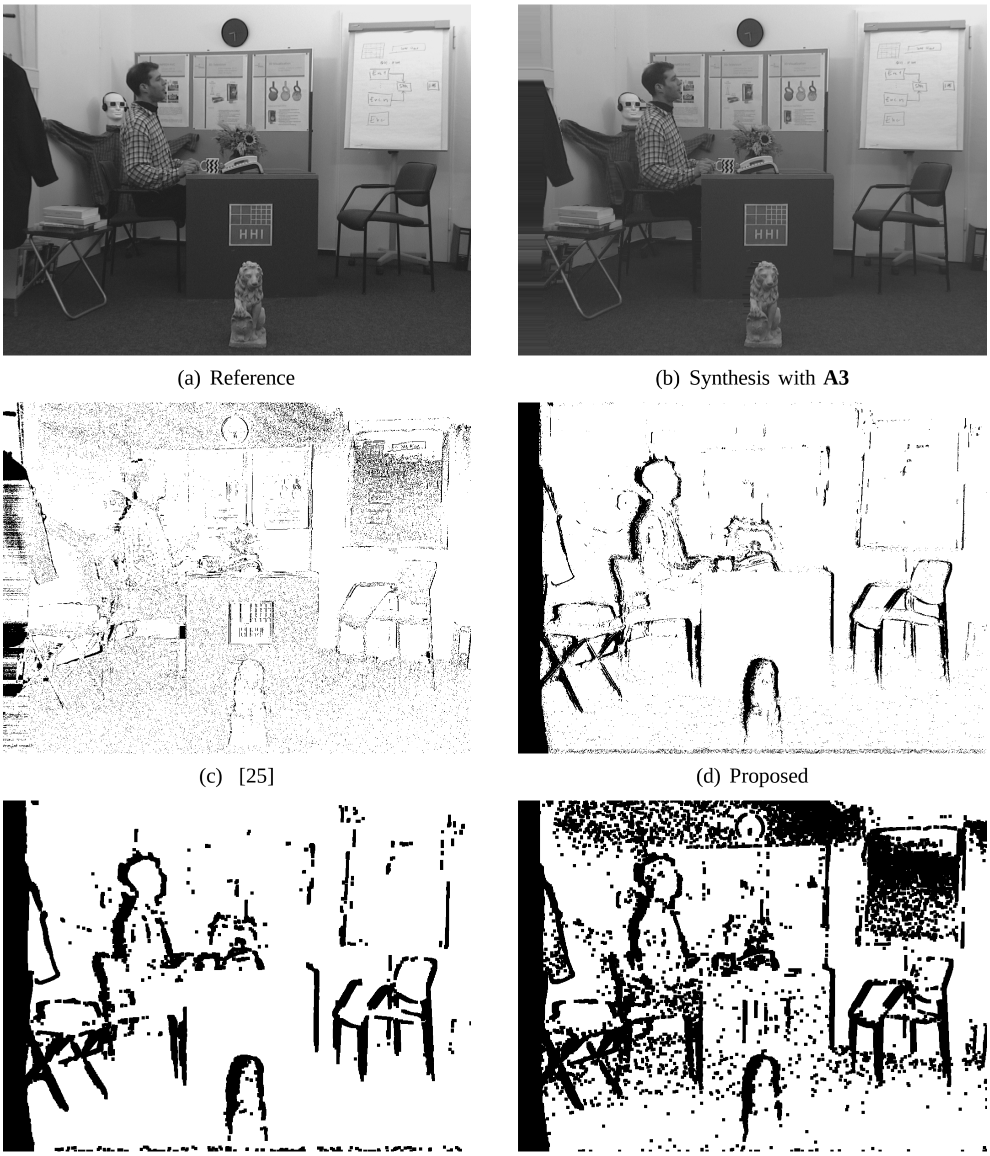
- **A1**: based on [30]. Depth map preprocessed by a low pass filter, borders are cropped and the image is resized to the original resolution.
- **A2**: based on [30] with inpainting algorithm proposed by Telea [31]
- **A3**: Tanimoto *et al.* [32], View Synthesis Reference Software (VSRS).
- **A4**: Muller *et al.* [33], depth aided inpainting
- **A5**: Ndjiki-Nya *et al.* [34], hole-filling using a patch-based texture synthesis.
- **A6**: Koppel *et al.* [35], synthesis is improved in disoccluded areas using depth temporal information
- **A7**: the disoccluded areas are not filled

Additional details on the database and an extensive study can be found in [36].

B. Testing methodology

In order to validate the results obtained with the proposed technique we want to evaluate all sequences and views, synthesized with each method. However, as the authors of [28] also notice there are some outliers in the methods. Method **A1** has the highest scores in the subjective tests while all objective metrics indicate this method is by far the worst. This is due to the method not using any inpainting algorithms to fill the disoccluded areas. The borders are cropped and the image is rescaled. The non-border disocclusions are avoided by performing a low-pass filtering of the depth map. While the final result is an image with no localized impactful artifacts, it cannot be used for 3D viewing, as the geometry of the scene no longer corresponds to the reference. These results also point out to the subjects inclination to notice localized artifacts more easily than a global change in the frame which further motivates the use of ROI evalution in synthesis methods. Since we analyze view synthesis for its capability of producing 3D content, we will not use this method in our results.

In our tests we use three quality evaluation metrics: Structural SIMilarity index (SSIM) [37], Peak-Signal-to-Noise-Ratio (PSNR) and Multi-scale SSIM (MSSIM) [38]. For each metric we apply the region of interest we described in Section II and the one proposed by Bosc *et al.* in [25]. For our method we use multiple variants: proposed mask (P) without erosion/dilation (e/d) or ground truth (GT); P with e/d and P with both e/d and GT. To measure the performance of each metric we compute the average values across frames for each sequence/view/method ($3 \times 4 \times 6$). In [25] the authors selected four critical points (subjective vs objective results) to evaluate the method. Our tests will be performed on all



(a) Reference (b) Synthesis with **A3**

(c) [25] (d) Proposed

(e) Proposed+e/d (f) Proposed+GT+e/d

Fig. 2. Book arrival sequence view 10 synthesized from view 8 with method **A3**. Luminance frames and binary masks for the proposed methods and [25]. Black pixels are selected for evaluation.

TABLE II. PCC, SROCC AND RMSE FOR NON-ROI, [25] AND OUR PROPOSED METHODS USING SSIM, PSNR AND MSSIM

Metric	Non-ROI			[2]			P			P+e/d			P+GT+e/d		
	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE
SSIM	60.85	49.94	47.16	61.29	58.64	47.08	**70.18**	**65.28**	**42.34**	69.00	56.63	43.02	68.88	55.46	43.09
PSNR	**85.97**	77.57	**30.36**	68.52	32.55	43.29	71.66	67.18	41.45	74.31	68.32	39.77	82.26	**79.20**	33.79
MSSIM	**80.10**	65.89	**35.58**	68.67	38.35	43.21	73.86	**70.69**	40.07	72.11	67.4	41.18	77.18	67.81	37.79

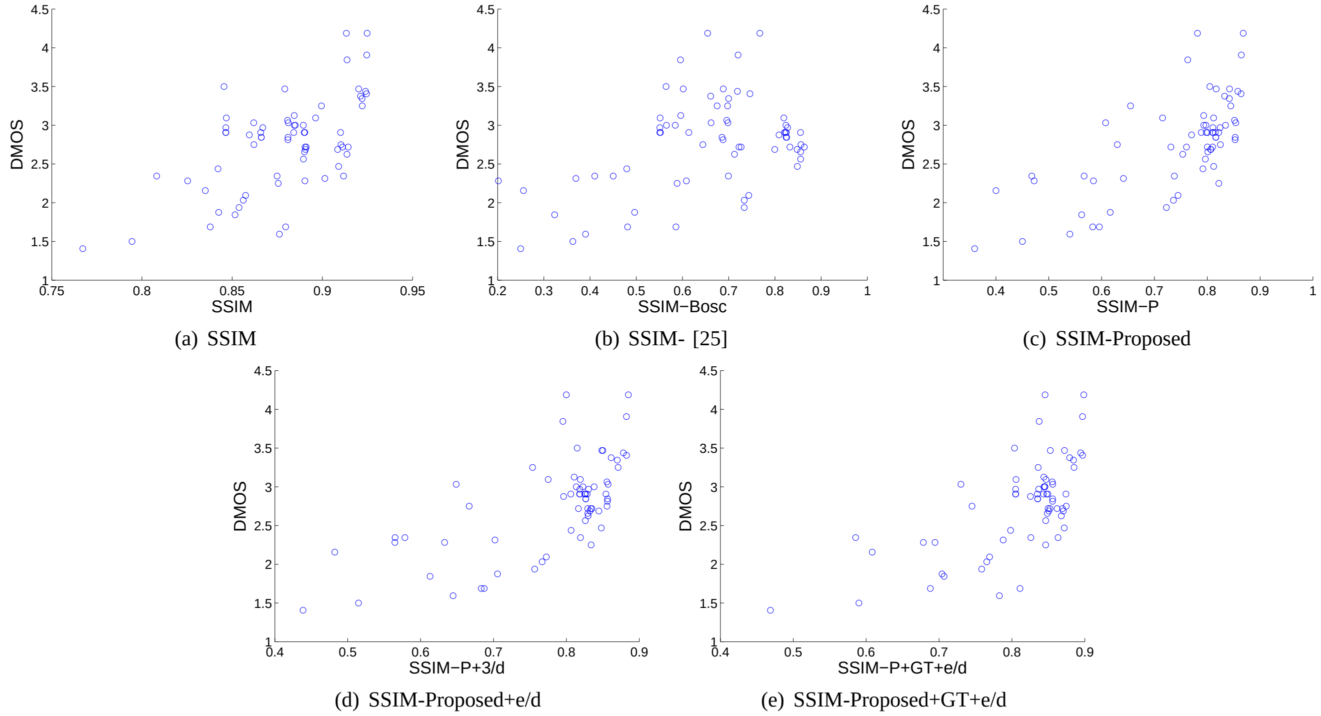
points using the Difference Mean Opinion Score (DMOS). The performance indicators we use are Pearson Correlation Coefficient (PCC), Spearman's Rank Order Correlation Coefficient (SROCC) and the Root-Mean-Squared-Error (RMSE). Before computing the PCC we will perform a fitting of the results using the recommended nonlinear function from VQEG Phase I final report [39]:

$$Y = \beta_2 + \frac{\beta_1 - \beta_2}{1 + \exp^{-\frac{X-\beta_3}{|\beta_4|}}} \tag{5}$$

where β_1, β_2, β_3, β_4 are parameters, Y are the predicted values and X are the objective results.

C. Results and discussion

In Figure 2 we show an example of generated masks for frame 10 of Book arrival sequence, view 10 synthesized from 8. Figures 2(a) and 2(b) show the reference and the synthesized frame with method **A3**. The filled dissoccluded areas are easy to notice on the left side of foreground objects



(a) SSIM (b) SSIM- [25] (c) SSIM-Proposed

(d) SSIM-Proposed+e/d (e) SSIM-Proposed+GT+e/d

Fig. 3. Scatter plots of objective results for SSIM with tested ROIs. Each point is the DMOS against the average objective score over all frames for a sequence, synthesis and method.

and also on the left border of the image. An additional source of errors which is harder to notice is also a slight displacement of certain textures on foreground object compared to the reference (e.g. the desk). Another source of errors is caused by a slight difference in luminance. This is common with DIBR synthesis methods. While they are able to warp objects to their new position in the virtual view, changes in luminance between views are not accounted. While this types of distortions are not visually impactful, as they are difficult to notice, they can have an impact on the results of objective metrics and are relevant to this study.

Figures 2(c), 2(d), 2(e) and 2(f) show the binary masks for [25], P, P+e/d and P+GT+e/d, respectively. When comparing 2(c) and 2(d) we can see that our mask is less noisy and better adjusted to the scene geometry. Also, the right side of the image, which corresponds to a border disocclusion is completely selected, as opposed to [25]. Furthermore, the texture details of the desk are not selected in our mask, because this area has a uniform depth and is rendered similarly with all DIBR methods. Although there is a slight displacement which will result in high errors, they are hard to notice and are not critical in differentiating the evaluated methods. Performing the e/d operation will reduce the isolated patches/pixels selected in the map while, increasing solid areas. Finally, adding the ground truth in the mask computation will lead to an increased selection. We can notice that additional textures are selected: the desk, the white board and the area surrounding the clock. In this example, the percentages of selected pixels are: 7.5%, 11.44%, 17.21%, 33.2% for Bosc [25], P, P+e/d and P+GT+e/d, respectively. This behavior is similar on other sequences/views/methods, however, for brevity reasons we only discuss this example.

In Figure 3 we show the scatter plots for SSIM and ROI SSIM with the binary masks [25], P, P+e/d and P+GT+e/d, respectively. Each point represents the DMOS against the average of the objective score over all frames of a sequence/view/method. An improvement can be observed when using our proposed approach. This is also reflected in the numerical results reported in Table II. Our methods outperforms [25] on all test cases. When compared to the Non-ROI scores, we are able to outperform SSIM with all proposed ROIs, while P+GT+e/d show similar performance to PSNR and MSSIM. A loss is observed with PSNR-P and MSSIM-P. This behavior can be explained by the use of e/d and GT. As discussed above the masks will have a larger number of selected pixels. Also, SSIM is already computed using a pixel's neighborhood, thus performing the e/d operation will allow PSNR-P+GT+e/d to account for the original's ROI neighborhood. However, the SSIM score will decrease in this case as pixels which are further away from the ROI are evaluated.

Another interesting aspect is the actual implementation for a ROI evaluation with different metrics. For MSSIM the tests were performed by rescaling the ROI. However, it is also possible to recompute the ROI using the rescaled images. Furthermore, additional metrics can be computed with respect to a ROI, though, in the case of perceptual based metrics the way to perform such an evaluation becomes more difficult.

IV. Conclusions

In this paper we presented a study on the use of ROI in the evaluation of DIBR based synthesis methods. We proposed a ROI generation method that can be used with traditional metrics, such as SSIM, PSNR and MSSIM. We validated this technique using a publicly available subjective evaluation database, for view synthesis methods, and showed that we can improve the objective results of SSIM, while maintaining similar results for PSNR and MSSIM when compared to subjective scores. Future directions may include finding a better threshold for the ROI selection by taking into account perceptual aspects or finding ways to use a ROI for perceptual metrics. Another study direction is to perform extensive subjective tests for view synthesis using more methods and also encoded reference views and depth maps.

References

[1] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "Free-Viewpoint TV," *IEEE Signal Processing Magazine*, vol. 28, pp. 67–76, 2011.

[2] F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, Eds., *Emerging technologies for 3D video: content creation, coding, transmission and rendering*. Wiley, May 2013.

[3] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," *IEEE International Conference on Image Processing*, vol. 1, pp. 201–204, 2007.

[4] C. Fehn, "A 3D-TV approach using depth-image-based rendering," in *3rd IASTED Conference on Visualization, Imaging, and Image Processing*, Benalmadena, Spain, 8-10 September 2003, pp. 482–487.

[5] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of SIGGRAPH*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42. [Online]. Available: http://doi.acm.org/10.1145/237170.237199

[6] H.-Y. Shum and L.-W. He, "Rendering with concentric mosaics," in *Proceedings SIGGRAPH*, Los Angeles, California USA, 1999, pp. 299–306.

[7] C. Buehler, M. Bosse, L. McMillan, and S. Gortler, "Unstructured Lumigraph Rendering," in *Proc SIGGRAPH*, Los Angeles, California USA, August 2001, pp. 425–432.

[8] F. Dufaux, M. Cagnazzo, and B. Pesquet-Popescu, *Motion Estimation - a Video Coding Viewpoint*, ser. Academic Press Library in Signal Processing, R. Chellappa and S. Theodoridis, Eds. Academic Press, 2014 (to be published), vol. 5: Image and Video Compression and Multimedia.

[9] L. Zhan-Wei, A. Ping, L. Su-xing, and Z. Zhao-yang, "Arbitrary view generation based on DIBR," in *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Xiamen, People's Republic of China, 2007, pp. 168–171.

[10] C.-M. Cheng, S.-J. Lin, S.-H. Lai, and J.-C. Yang, "Improved novel view synthesis from depth image with large baseline," in *19th International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[11] "High Efficiency Video Coding," ITU-T Recommendation H.265 and ISO/IEC 23008-2 HEVC, April 2013.

[12] "Report on experimental framework for 3D video coding," ISO/IEC JTC1/SC29/WG11 MPEG2010/N11631, October 2010.

[13] L. Zhang, G. Tech, K. Wegner, and S. Yea, "3D-HEVC test model 5," ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 JCT3V-E1005, July 2013.

[14] S. Huq, A. Koschan, and M. Abidi, "Occlusion filling in stereo: theory and experiments," *Computer Vision and Image Understanding*, vol. 117, pp. 688–704, June 2013.

[15] I. Daribo and B. Pesquet-Popescu, "Depth-aided image inpainting for novel view synthesis," in *IEEE MMSP*, Saint Malo, France, 4-6, October 2010.

[16] C. Guillemot and O. L. Meur, "Image inpainting: Overview and recent advances," *IEEE Signal Processing Magazine*, vol. 31, pp. 127–144, 2014.

[17] A. Criminisi, P. Perez, 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.

[18] P.-J. Lee and Effendi, "Adaptive edge-oriented depth image smoothing approach for depth image based rendering," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Shanghai, 24-26 March, 2010, pp. 1–5.

[19] Z. Wang and J. Zhou, "A novel approach for depth image based rendering, based on non-linear transformation of depth values," in *International Conference on Image Analysis and Signal Processing (IASP)*, Hubei, People's Republic of China, 21-23 October 2011, pp. 138–142.

[20] A. Purica, E. G. M., B. Pesquet-Popescu, M. Cagnazzo, and B. Ionescu, "Improved view synthesis by motion warping and temporal hole filling," in *ICASSP*. South Brisbane: IEEE, 19-24 April 2014, pp. 1191–1195.

[21] W. Sun, O. C. Au, L. Xu, Y. Li, and W. Hu, "Novel temporal domain hole filling based on background modeling for view synthesis," in *IEEE International on Image Processing (ICIP)*, Orlando, FL, 30 Sept. - 3 Oct. 2012, pp. 2721 – 2724.

[22] K. P. Kumar, S. Gupta, and K. S. Venkatesh, "Spatio-temporal multi-view synthesis for free viewpoint television," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, Aberdeen, 7-8 October 2013, pp. 1 – 4.

[23] VQEG 3DTV Group. http://www.its.bldrdoc.gov/vqeg/projects/3dtv/.

[24] A. Tikanamaki, A. Gotchev, and A. S. S. Miller, "Quality assessment of 3-d video in rate allocation experiments," in *IEEE International Symposium on Consumer Electronics, 2008. ISCE 2008*. Vilamoura: IEEE, 14-16 April 2008, pp. 1–4.

[25] E. Bosc, R. Pepion, P. L. 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, September 2011.

[26] A. Purica, M. Cagnazzo, B. Pesquet-Popescu, F. Dufaux, and B. Ionescu, "A distortion evaluation framework in 3d video view synthesis," in *International Conference on 3D Imaging (IC3D)*. Liege, Belgium: IEEE, 14-15 December 2015, pp. 1–8.

[27] S. L. P. Yasakethu, C. Hewage, W. Fernando, and A. Kondoz, "Quality analysis for 3-d video using 2-d video quality models," *IEEE Trans. Consumer Electron.*, vol. 54, no. 4, pp. 1969–1976, November 2008.

[28] DIBR videos quality assessment (using acr-hr). [Online]. Available: http://ivc.univ-nantes.fr/en/databases/DIBR_Videos/

[29] "Itu-t study group 12," ITU-T p.910 Subjective Video Quality Assessment Methods for Multimedia Applications, 1997.

[30] C. Fehn, "Depth-image-based-rendering (dibr) , compression and transmission for a new approach on 3D-TV," in *Proc. of SPIE Stereoscopic Displays and Virtual Reality Sistems*, vol. 5291, 2004, pp. 93–104.

[31] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics, GPU and Game Tools*, vol. 9, pp. 25–36, 2004.

[32] M. Tanimoto, T. Fujii, K.Suzuki, N. Fukushima, and Y. Mori. Reference software for depth estimation and view synthesis. ISO/IEC JTC1/SC29/WG11 MPEG 2008/M15377. MPEG.

[33] K. Muller, A. Smolic, K. Dix, P. Merkle, and P. Kauff, "View synthesis for advanced 3d video systems," *EURASIP Journal on Image and Video Processing*, 2008.

[34] 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, June 2011.

[35] M. Koppel, P. Ndjiki-Nya, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," in *17th IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 1809–1812.

[36] E. Bosc, P. Le Callet, L. Morin, and M. Pressigout, *3D-TV System with Depth-Image-Based Rendering Architectures, Techniques and Challenges*. Springer, 2012.

[37] 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, pp. 600–612, 2004.

[38] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, November 2004, pp. 1398–1402.

[39] Video Quality Experts Group. (2000, March) Final report from the video quality experts group on the validation of objective models of video quality assessment. VQEG.