

AN EDGE-BASED STRUCTURAL DISTORTION INDICATOR FOR THE QUALITY ASSESSMENT OF 3D SYNTHESIZED VIEWS

Emilie Bosc¹, Patrick Le Callet², Luce Morin¹, Muriel Pressigout¹

¹ IETR UMR CNRS 6164, INSA de Rennes

20, avenue des Buttes de Coesmes, 35708 RENNES CEDEX 7, France

² LUNAM Université, Université de Nantes, IRCCyN UMR CNRS 6597

Polytech Nantes, rue Christian Pauc BP 50609 44306 Nantes Cedex 3, France

ABSTRACT

3D-TV applications require the generation of novel viewpoints through Depth-Image-Based-Rendering methods. These synthesized views need to be assessed by a reliable quality metric. Most of the proposed metrics are inspired from 2D commonly used quality metrics. Yet, the latter were originally designed to address 2D compression distortions which are different from the distortions related to DIBR processes. We propose an edge-based method that indicates the level of structural degradation in the synthesized image. The first results are encouraging since the correlation to subjective scores is higher than other tested metrics.

Index Terms— Virtual view synthesis, DIBR, multi-view video, 3DTV, quality assessment, quality metric

I. INTRODUCTION

Multi-view plus Depth data represents a scene through a set of color sequences, acquired at slightly different viewpoints, and a set of associated depth sequences. From depth and color information a novel view (also called a synthesized view) can be generated thanks to a Depth-Image-Based-Rendering (DIBR) method [1]. The major issue in DIBR systems consists in filling in the disoccluded regions of the novel viewpoint: when generating a novel viewpoint, regions that were not visible in the former viewpoint, become visible in the novel viewpoint. However, the appropriate color information related to these discovered regions is often unknown. In-painting methods that are either extrapolation or interpolation techniques, are meant to fill the disoccluded regions. However, distortions from in-painting are specific and dependent on a given technique, as observed in [2]. Moreover, due to the rounding of pixel positions when projecting the color information in the target viewpoint, the pixels mapped in the target viewpoint may not locate at an integer position. In this case the position is either rounded to the nearest integer or interpolated. Finally, another source of distortion relies on the depth map uncertainties. Errors in depth maps estimation cause visual distortion in the synthesized views because the color pixels are not correctly



Fig. 2. Synthesized views - Frame 45 - view 10 - *Book Arrival*.

mapped. As well, the problem is similar when depth maps suffer important quantization from compression methods [3]. Fig. 2 and Fig. 3 give examples of distortions. The synthesized views depicted in these figures were obtained through different DIBR algorithms. As it can be observed, the distortions are located around the edges of the arms for Fig. 2 and around the edges of the face for Fig. 3. This corresponds to strong depth discontinuities. Thus, artefacts related to DIBR systems are mostly located in specific areas that are the disoccluded regions. They are not scattered in the entire image such as specific 2-D video compression distortions. Consequently, as shown in [2], commonly used 2-D quality metrics, that were originally design to address 2-D video compression distortions, are not sufficient enough to assess the visual quality of synthesized views. They indicate the presence of errors but not the degree of visual annoyance. A novel assessment tool can either improve the existing 2-D metrics, or propose a new approach.

Yasakethu *et al.* [4] proposed an adapted VQM for measuring 3-D Video quality. It combines 2-D color information quality and depth information quality. Depth quality measurement is based on an analysis of the depth planes distortion. Results show higher correlation to subjective scores than simple VQM. Another approach improving the existing 2-D metrics is the method proposed by Ekmekcioglu *et al.* [5]. This depth-based perceptual tool can be applied

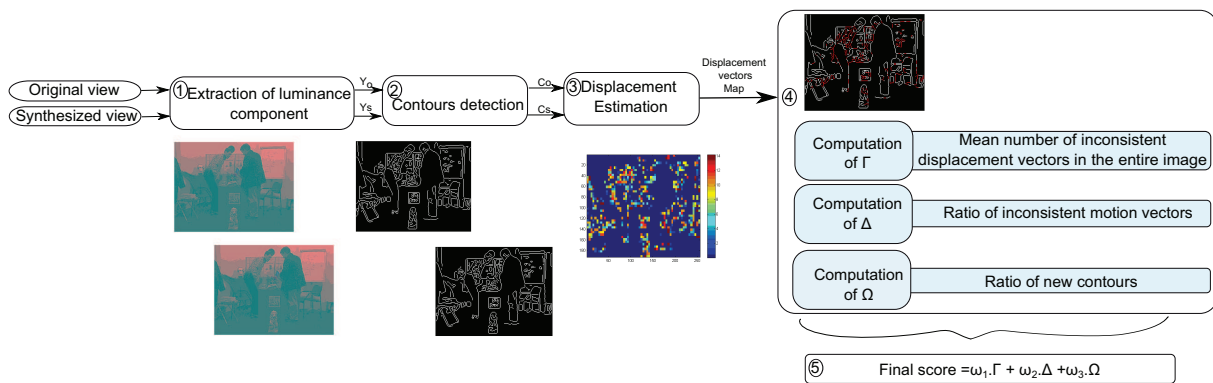


Fig. 1. Overview of the proposed method

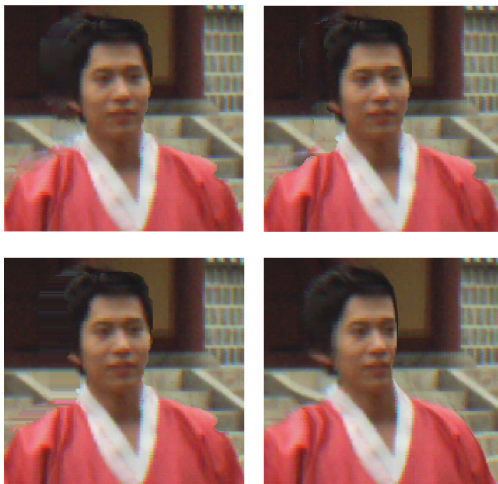


Fig. 3. Synthesized views - Frame 112 - view 6 - *Lovebird1*.

to PSNR or SSIM. The method uses a weighting function based on depth data at the target viewpoint, and a temporal consistency function to take the motion activity into account. The study [5] shows that the proposed method enhances the correlation of PSNR and SSIM to subjective scores.

Our proposed method is based on the analysis of the synthesized view edges compared to the edges of the original acquired view. The paper is organized as follows. Section II presents the principles of the proposed approach. Section III presents the experiments, the validation of the method and its limitations. Finally Section IV concludes the paper.

II. PROPOSED INDICATOR

Our proposed method is based on the observation of the nature and of the location of the artefacts, as pointed out in [2]. The method relies on the analysis of the synthesized view edges in comparison to the original image edges.

More than a quality measure it is an indicator of structural distortion. Let I_o be the original view, I_s is the synthesized view. As depicted in Fig. 1, first step consists in extracting the luminance component of I_o and I_s , referred as Y_o and Y_s respectively. At the second step a Canny edge detector is applied on Y_o and Y_s . Let C_o and C_s be the resulting extracted contours. At the third step a displacement vector estimation is performed between C_o and C_s . The resulting displacement vectors map is processed at the fourth step. Three parameters are computed: the mean ratio of inconsistent displacement vectors per contour pixel Γ , the ratio of inconsistent vectors Δ , the ratio of new contours Ω . They are defined as follows:

$$\Gamma = \frac{1}{|C_s|} \sum_{c=1}^{|C_s|} \gamma_c \quad (1)$$

where γ_c is the ratio of inconsistent displacement vectors for the pixel $C_s(c)$. It is defined as:

$$\gamma_c = \frac{1}{K} \sum_n \delta(i, n) \quad (2)$$

with N the size of the slide window, i and n are such as $Y_s(i) \in C_s$ and $Y_s(n) \in C_s$, K is a normalizing factor, and

$$\delta(i, n) = \begin{cases} 1, & \text{if } \overrightarrow{M_i M_n} > Th \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

M_i and M_n are the displacement vectors of $Y_s(i)$ and $Y_s(n)$, $\overrightarrow{M_i M_n}$ is the angle formed between M_i and M_n , and Th is a threshold. The final score is a weighting sum of the three parameters:

$$Indicator = 1 - (\alpha_1 \Gamma + \alpha_2 \Delta + \alpha_3 \Omega) \quad (4)$$

In the experiments we used the combination $\{Th = 45^\circ, \alpha_1 = 0.25, \alpha_2 = 0.25, \alpha_3 = 0.5\}$. The closer the final score is to 1, the less distorted are the contours of the image.

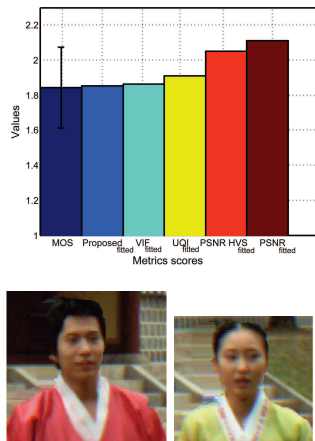


Fig. 4. Quality evaluation of a synthesized view (Frame 112 - view 6 - *Lovebird1* rendered with [9])

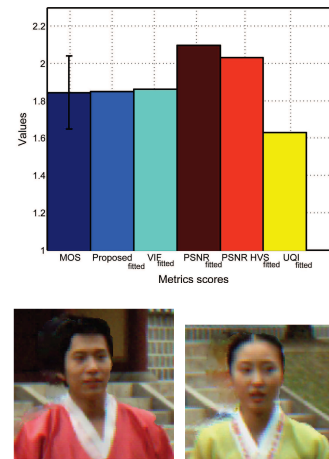


Fig. 5. Quality evaluation of a synthesized view (Frame 112 - view 6 - *Lovebird1* rendered with [10])

III. EXPERIMENTAL RESULTS AND DISCUSSION

The quality of the set of synthesized views used in [2] is assessed through commonly used metrics and through the proposed indicator. The obtained scores are fitted and scaled into a common MOS scale. Fig. 4, 5, 6 and 7 depict the quality scores of four synthesized views containing different type of distortions (color leak, blurry regions, and geometric distortions). In each figure, a bar plot gives the Mean Opinion Score (MOS), the PSNR fitted score, the fitted score the closest to the MOS score among the objective metrics fitted scores, and the proposed indicator fitted score. The objective metric whose fitted score was closest to the MOS was not the same for all the figure, but we provide it for both figure. For example, in Fig. 4 the objective metric whose fitted score is the closest to the MOS is VIF, but in Fig. 6 it is PSNR-HVS. Also, particular regions of the synthesized views are provided in each figure. In the presented cases, PSNR shows the highest gap with the MOS score: even if a distortion is not perceptible, it contributes to the decrease of the quality score because its perceptual impact on the quality is not considered. The objective score that is the closest to the MOS is also provided. Although the closest objective score is different depending on the figure, it can be observed that it is a Human Visual System (HVS) based metric in both of the presented cases (Visual Information Fidelity (VIF) [6], PSNR-HVS [7] and Universal Quality Index [8] (UQI)). Contrary to the tested objective metrics, in both of the presented cases, our proposed indicator was close to the MOS or to the closest.

However, the proposed method does not assess the image quality, but it is able to detect the structural distortions in this context. These results are very encouraging because at this stage the proposed method can only assess structural distortions. To be considered as a quality metric, the color

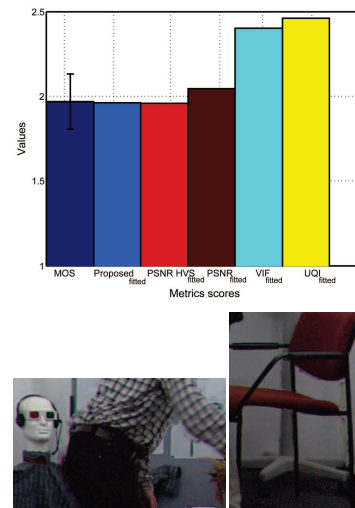


Fig. 6. Quality evaluation of a synthesized view (Frame 54 - view 10 - *Book Arrival* rendered with [10])

consistency should also be analyzed and assessed. Moreover, the influence of weights α_i in Eq. (4) must be studied in future works.

IV. CONCLUSION

We proposed a edged-based structural distortion indicator to address the specific distortions related to DIBR systems. The results are encouraging because the proposed method improved the prediction of human judgment. However, since it does not take the color consistency into account, the proposed method remains a tool for assessing the structural consistency of an image. Further investigation will be focused on the color consistency.

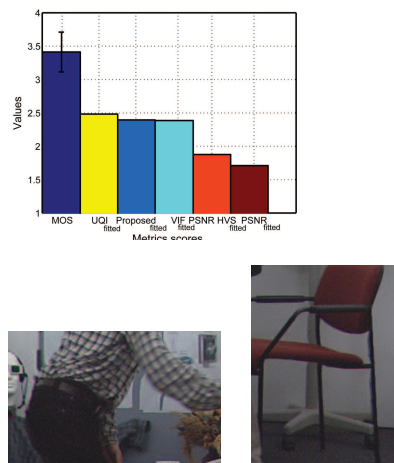


Fig. 7. Quality evaluation of a synthesized view (Frame 54 - view 9 - *Book Arrival* rendered with [1])

V. ACKNOWLEDGMENTS

This work is partly supported by the French ANR-PERSEE project n° ANR-09-BLAN-0170, and ANR-CAIMAN project n° ANR-08-VERS-002.

VI. REFERENCES

- [1] C. Fehn, "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*, 2004, vol. 5291, p. 93104.
- [2] E. Bosc, R. Pepion, P. Le Callet, M. Koppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, "Towards a new quality metric for 3D synthesized view assessment," *IEEE Journal of Selected Topics*, 2011.
- [3] D. V.S.X De Silva, W. A.C Fernando, and S. T Worrall, "Intra mode selection method for depth maps of 3D video based on rendering distortion modeling," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2735–2740, Nov. 2010.
- [4] S. L. P. Yasakethu, S. T Worrall, D. De Silva, W. A. C. Fernando, and A. M. Kondoz, "A compound depth and image quality metric for measuring the effects of packet loss on 3D video," in *Proc. of 17th International Conference on Digital Signal Processing*, Corfu, Greece, July 2011.
- [5] E. Ekmekcioglu, S. T. Worrall, D. De Silva, W. A. C. Fernando, and A. M. Kondoz, "Depth based perceptual quality assessment for synthesized camera viewpoints," in *Proc. of Second International Conference on User Centric Media, UCMedia 2010*, Palma de Mallorca, Sept. 2010.
- [6] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430444, 2006.
- [7] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS," in *CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, USA, 2006.
- [8] Z. Wang and A. C. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81–84, 2002.
- [9] 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*, 2008.
- [10] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, "Reference softwares for depth estimation and view synthesis," Apr. 2008.