https://doi.org/10.2352/ISSN.2470-1173.2018.12.IQSP-366

A full-reference Image Quality Assessment metric for 3D Synthesized Views

Shishun TIAN, Lu ZHANG, Luce MORIN, Olivier DEFORGES; IETR UMR-CNRS 6164; INSA de Rennes; Rennes, France

Abstract

The quality assessment of Depth-Image-Based-Rendering (DIBR) synthesized views is very challenging owing to the new types of distortions, thus the traditional 2D quality metrics may fail to evaluate the quality of the synthesized views. In this paper, we propose a full-reference metric to assess the quality of DIBR synthesized views. Firstly, we notice that the object shift in the synthesized view is approximately linear, an affine transformation is used to warp the pixel in the reference image to the corresponding position in the distorted image. Besides, since the synthesis distortions mainly happen in the dis-occluded areas, a dis-occlusion mask obtained from the depth map in the original viewpoint is used to weight the final distortions between the synthesized image and the reference image. The experimental results on IRCCyN/IVC DIBR image database show that the proposed weighted PSNR (PSNR′) outperforms the state-of-the-art DIBR synthesized view dedicated metrics: 3DSwIM, VSQA, MP-PSNR, MW-PSNR and earns a gain of 36.85% (in terms of PLCC) over PSNR. The weighted SSIM (SSIM′) earns a gain of 13.33% (in terms of PLCC) compared to SSIM.

Introduction

Providing more immersive perception of a visual scene, 3D video applications, such as 3D-TV [1] and Free-viewpoint TV (FTV) [2], have gained great public interest and curiosity in recent years. In most 3D applications, 3D content is obtained by using multiple cameras to record the same scene at slightly different viewpoints. Multiview-Video-Plus-Depth (MVD) [3] format is one of the most widely used 3D representations. It consists of texture views and their associated depth maps from different viewpoints. Furthermore, this 3D representation can be exploited by generating virtual viewpoints, thus the virtual view from a viewpoint which has not been recorded can be rendered by using Depth-Image-Based-Rendering (DIBR). MVD and DIBR can be used for many 3D applications, such as FTV which is able to allow the users to view a 3D scene by freely changing their viewpoints.

Benefiting from DIBR, only a limited number of original views needs to be coded and transferred, the additional virtual views on the receiver side can be synthesized from the decoded views. However, this process will lead to some new kind of distortions due to depth errors, occlusions and inpainting methods. These distortions are quite different from the ones induced by 2D image compression, since most video coding standards rely on Discrete Cosine Transform, which leads to specific artifacts [4]. These distortions are often scattered over the whole image, whereas the Depth-Image-Based-Rendering (DIBR) synthesized artifacts which are mainly caused by depth compression and view synthesis usually happen in the dis-occluded areas. Besides, inaccurate depth maps can also introduce various distortions, such as object shifting and geometric distortions in the synthesized views. Since most of the objective quality metrics were initially designed to assess specific usual distortions, they may fail in assessing the quality of images containing view synthesis distortions [5][6]. Meanwhile, the use of subjective tests is time consuming and practically not suitable for those applications where a real-time quality score is needed. Efficient objective metrics are thus urgently needed to assess the quality of synthesized views.

In the litterature, several full-reference (FR) methods have been proposed to assess the quality of synthesized images, such as View Synthesis Quality Assessment (VSQA) [7], 3D Synthesized view Image Quality Metric (3DSwIM) [8], Morphological Wavelet Peak Signal-to-Noise Ratio measure (MW-PSNR) [9] and Morphological Pyramid Peak Signal-to-Noise Ratio (MP-PSNR) [10]. The principle of VSQA [7] is to apply three weighting maps on the SSIM distortion map [11] to characterize the image complexity in terms of textures, diversity of gradient orientations and presence of high contrast. It is reported that it approaches a gain of 17.8% over SSIM in the correlation with subjective measurement. In [8], Battisti et al. proposed 3DSwIM, a metric based on a comparison of statistical features of wavelet sub-bands of the original and DIBR-synthesized images. Only horizontal detail sub-bands are utilized since the synthesis artifacts mainly happen in the horizontal direction. A registration and a skin detection process are also used to make sure that the best matching blocks and the most sensitive impairments are always compared. 3DSwIM is reported to outperform the traditional 2D metrics and existing DIBR image dedicated metrics. Sandic-Stankovic et al. proposed a morphological wavelet decomposition based metric MW-PSNR[9] and a morphological pyramid decomposition based metric MP-PSNR [10]. The non-linear morphological filters are used to maintain important geometric information such as edges across different resolution levels. Besides, in [12], the same authors have also proposed the reduced version (MP-PSNRr and MWPSNRr) of MP-PSNR, and MW-PSNR, which only use detail images from higher decomposition scales. The experimental results show that they achieve higher correlation with human judgment compared to the state-of-art image quality assessment metrics.

In this paper, we propose a full-reference quality assessment metric for 3D synthesized views by compensating the object shift and using a disparity map as a mask to weight the final distortion. In the following section, we detail the proposed method. In the third section, we present and discuss the experimental results. Finally, we conclude the paper in the last section.

Proposed method

In this section, we propose a full-reference metric for DIBR synthesized views. First of all, we compensate the global shifting caused by resizing in some synthesis algorithms according to the fact that it will be punished by pixel-based metrics while is not perceivable by human observers. Besides, as the synthesis distortions mainly occur in the dis-occluded areas, a dis-occluded mask is obtained to weight the final distortions.

Shift Compensation

As shown in Fig.1, we observed that global shifting mainly occurs in the horizontal direction and this shift could be recognized as approximately linear. It is modeled by an $affine$ transformation [13].

Firstly, SURF feature detection [14] is utilized to detect the feature points in the reference and synthesized view. The RANSAC algorithm [15] is used to estimate the matrix H associated with the $affine$ transform. Then, the synthesized view is warped to the reference view by the obtained transformation matrix H. The optimized matched feature point pairs are shown in Fig. 1. The *SSIM maps* before and after shift compensation are shown in Fig. 2.

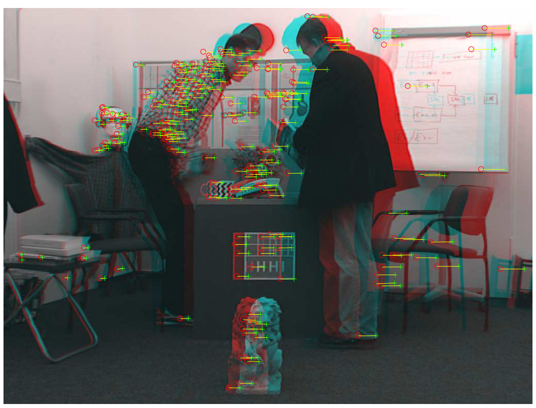
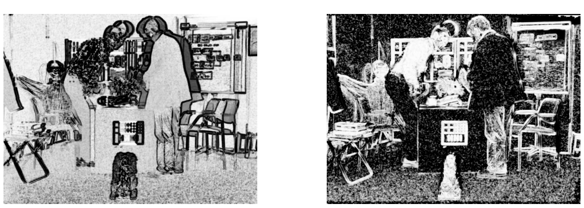


Figure 1. *Example of optimized matched feature point pairs*



(a) SSIM map before transform (b) SSIM map after transform

Figure 2. *Example of SSIM maps before and after the transformation*

Dis-occlusion Mask

Since the synthesis distortion mainly occurs in the dis-occluded areas, we utilize a dis-occlusion mask to weight the final distortion. The depth map in the original view-point ($Depth_o$) is used to calculate the dis-occlusion mask. In a matched configuration 3D warping process, the horizontal disparity which is the horizontal displacement for each pixel can be obtained by Eq. (1):

$$d = \frac{f \times l}{Z} \quad (1)$$

where f, l, Z represent the camera focal length, the baseline distance between these two views and the depth value of this pixel respectively.

The depth map in the synthesized view-point ($Depth_s$) given initial value to -1, then the depth map in the original view-point ($Depth_o$) is warped to the synthesized view-point by Eq. (2):

$$Depth_s(i+d,:) = Depth_o(i,:); \quad (i+d), i \in [1,W] \quad (2)$$

where W is the image width.

The dis-occluded mask dis_mask can then be obtained by extracting all the pixels with value -1 in $Depth_s$, which is shown in Fig. 3.

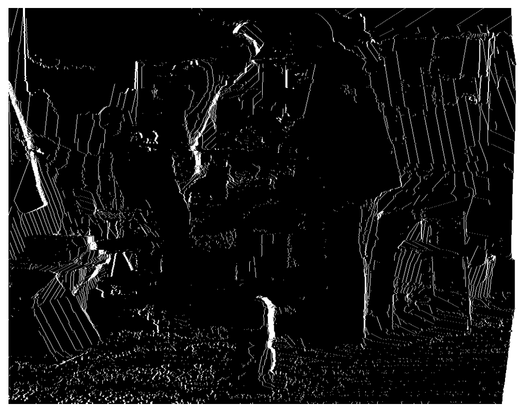


Figure 3. *Example of dis-occluded mask*

Weighted PSNR and Weighted SSIM

Generally speaking, the dis-occlusion mask dis_mask, can be integrated into any existing full-reference metric as a weighting mask. In this paper, we propose and test the weighted $PSNR$ ($PSNR'$) and $SSIM$ ($SSIM'$) as defined in the following equations:

$$MSE' = \frac{\sum_{(i,j)\in I}(I_{syn}(i,j) - I_{ref}(i,j))^2 \cdot dis\_mask(i,j)}{\sum_{(i,j)\in I} dis\_mask(i,j)} \quad (3)$$

$$PSNR' = 10 \cdot log_{10}(\frac{255 \times 255}{MSE'}) \quad (4)$$

$$SSIM' = \frac{\sum_{(i,j)\in I} SSIM(i,j) \cdot dis\_mask(i,j)}{\sum_{(i,j)\in I} dis\_mask(i,j)} \quad (5)$$

where I_{syn} and I_{ref} denote the the compensated synthesized image and the reference image respectively; dis_mask denotes the obtained disparity mask; $SSIM$ denotes the $SSIM$ map between the compensated synthesized image and the reference image.

Experimental results

The performance of the proposed method is evaluated on the IRCCyN/IVC DIBR database [16][17]. It contains the images generated from three different multi-view plus depth (MVD) sequences : Book Arrival(1024×768, 16 cameras with 6.5 cm spac-ing), Lovebird1(1024×768, 12 cameras with 3.5 cm spac-ing) and Newspaper(1024×768, 9 cameras with 5 cm spac-ing). For each sequence, four virtual views are generated using seven different DIBR synthesis algorithms A1-A7 [18, 19, 20, 21, 22]. This database consists of 84 synthesized views and their associated 12 original views along with subjective score - mean opinion score (MOS). Usually, Differential Mean Opinion Score (DMOS) is used as subjective score. In this paper, the DMOS is obtained via Eq. 6 [23].

$$DMOS = MOS_{syn} - MOS_{ref} + 5 \tag{6}$$

where MOS_{syn} and MOS_{ref} represent the MOSs of the synthesized image and the reference image respectively.

In order to compare the performance between the proposed metric and the state-of-the-art DIBR dedicated metrics, the following 3 widely employed crieria are used: Pearson Linear Correlation Coefficients (PLCC), Spearman's Rank Order Correlation Coefficients (SROCC) and Root-Mean-Square-Error (RMSE). Before calculating these 3 figures of metric, the objective scores need to be fitted to the predicted DMOS ($DMOS_p$) using Eq. 7, as recommended by Video Quality Expert Group (VQEG) Phase I FR-TV [23].

$$DMOS_p = a \cdot scores^3 + b \cdot score^2 + c \cdot score + d \tag{7}$$

where $score$ is the score obtained by the objective metric and a, b, c, d are the parameters of the cubic function. They are obtained through regression to minimize the difference between $DMOS_p$ and $DMOS$. The scatter plot of $DMOS$ versus the proposed weighted $PSNR$ and $SSIM$ are shown in Fig. 4 and Fig. 5.

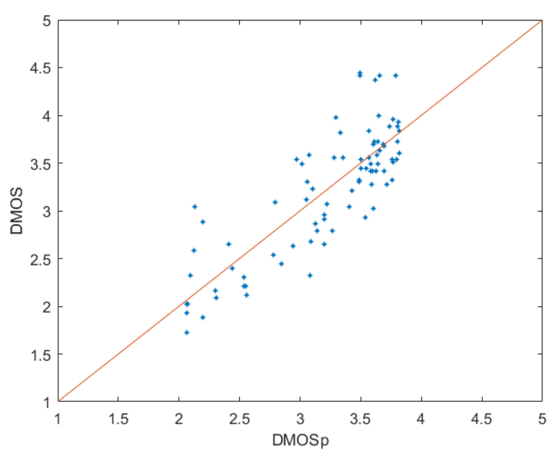


Figure 4. *Scatter plot of DMOS versus the proposed weighted PSNR*

We conduct performance comparison between the proposed method and two commonly used 2D FR metrics: PSNR, SSIM; and six state-of-the-art synthesized view dedicated metrics: MP-PSNR, MW-PSNR, MP-PSNRr, MW-PSNRr, 3DSwIM, VSQA. Their PLCC, RMSE, SROCC results are shown in Table 1. It can

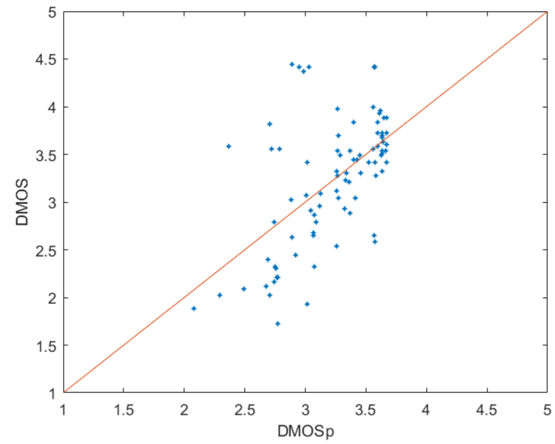


Figure 5. *Scatter plot of DMOS versus the proposed weighted SSIM*

Performance comparison of the proposed method with the state-of-the-art metrics

Metric	PLCC	RMSE	SROCC
PSNR DMOSp	0.4557	0.5927	0.4417
SSIM	0.4348	0.5996	0.4004
3DSwIM	0.6864	0.4842	0.6125
VSQA	0.6122	0.5265	0.6032
MP-PSNR	0.6729	0.4925	0.6272
MW-PSNR	0.6200	0.5224	0.5739
MP-PSNRr	0.6954	0.4784	0.6606
MW-PSNRr	0.6625	0.4987	0.6232
PSNR'(pro)	0.8242	0.3771	0.7889
SSIM'(pro)	0.5681	0.5479	0.5475

be noticed that the proposed weighted PSNR ($PSNR'$) performs the best among tested metrics. Its gain of PLCC achieves 36.85% compared to the PSNR. The weighted SSIM ($SSIM'$) achieves a gain of PLCC 13.33% compared to the SSIM.

Conclusion

In this paper, we proposed a full-reference quality metric dedicated for 3D synthesized views. The great advantage is its simplicity. The idea is to improve the existing simple 2D metrics by addressing two issues: 1) compensating the global significant shift in the synthesized view (by an affine transformation here); 2) putting more weights on the distortions occurring in the dis-occluded regions (which are estimated using the depth map here). Experimental results show that the proposed weighted PSNR ($PSNR'$) greatly improves the performance compared to the original PSNR (gain of 36.85% in terms of PLCC) and outperforms the tested state-of-the-art 3D synthesized view dedicated metrics: 3DSwIM, MP-PSNR, MW-PSNR. The weighted SSIM ($SSIM'$) earns also a gain of 13.33% (PLCC) compared to its 2D version. As future work, we plan to compensate the slight shift in a more precise way since we only compensate the global shift in this work. Besides, as saliency map can reflect the visual attention, another interesting way to explore is to improve the proposed method by combing it with a saliency detection method.

References

[1] C. Fehn, "A 3D-TV approach using depth-image-based rendering (DIBR)," in *Proc. of VIIP*, vol. 3, no. 3, 2003.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] F. Battisti, E. Bosc, M. Carli, P. Le Callet, and S. Perugia, "Objective image quality assessment of 3d synthesized views," *Signal Processing: Image Communication*, vol. 30, pp. 78–88, 2015.

[9] D. Sandić-Stanković, D. Kukolj, 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.

[10] D. Sandić-Stanković, D. Kukolj, 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.

[11] 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.

[12] D. Sandić-Stanković, D. Kukolj, 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.

[13] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *IEEE transactions on medical imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[14] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Computer vision–ECCV 2006*, pp. 404–417, 2006.

[15] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[16] IVC-IRCCyN lab, "IRCCyN/IVC DIBR image database," http://ivc.univ-nantes.fr/en/databases/DIBR_Images/.

[17] 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.

[18] 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.

[19] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3d warping using depth information for ftv," *Signal Processing: Image Communication*, vol. 24, no. 1, pp. 65–72, 2009.

[20] 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.

[21] 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.

[22] 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.

[23] Video Quality Experts Group, "Final report from the video quality experts group on the validation of objective models of multimedia quality assessment," *VQEG*, March 2008.

Author Biography

Shishun TIAN received his B.S. degree in electronics engineering from College of Electronics and Information Engineering, Sichuan University (2012) and his M.S. degree in optical engineering from University of Chinese Academy of Sciences (2015). He is currently pursuing his Ph.D degree from National Institute of Applied Sciences (INSA), Rennes, France, and in the Institute of Electronics and Telecommunications of Rennes (IETR) Laboratory. His research interests include image quality assessment and visual perception.

Lu ZHANG received the B.S. degree from Southeast University, China (2004), the M.S. degree from Shanghai Jiaotong University, China (2007) and her PhD degree from the LISA (now named LARIS) and CNRS IRCCyN (now named LS2N) labs in France (2012) respectively. She is currently an associate professor at National Institute of Applied Sciences (INSA), and the IETR lab, Rennes, France. She is now working on natural and medical image quality assessment, human perception understanding, image saliency detection, etc.

Luce Morin is Full-Professor at the Electrical and Computer Engineering Department in the National Institute of Applied Sciences (INSA) and a researcher at the Institute of Electronics and Telecommunications of Rennes (IETR). She leads the VAADER

team in the IETR laboratory. She has published more than 70 scientific papers in international journals and conferences. Her research activities deal with computer vision, 3D reconstruction, image and video compression, and representations for 3D videos and multiview videos.

Olivier Dforges received the Ph.D. degree in image processing from Polytechnique Nantes, France, in 1995. He has been involved with the ISO MPEG Standardization Group since 2007. He is currently a Full-Professor with Institut National des Sciences Appliques de Rennes (INSA) and IETR lab, Rennes, France. He has authored more than 180 technical papers. His research interests include image and video lossy and lossless compression, image understanding, fast prototyping, and parallel architectures.