QUALITY ASSESSMENT OF 3D SYNTHESIZED IMAGES VIA DISOCCLUDED REGION DISCOVERY

Yu Zhou[†], *Leida Li*[†], *Ke Gu*[‡], *Yuming Fang*[⋆] *and Weisi Lin*[‡]

[†]School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, 221116, China
[‡]School of Computer Engineering, Nanyang Technological University, 639798, Singapore
[⋆]School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, 330032, China

ABSTRACT

Depth-Image-Based-Rendering (DIBR) is fundamental in free-viewpoint 3D video, which has been widely used to generate synthesized views from multi-view images. The majority of DIBR algorithms cause disoccluded regions, which are the areas invisible in original views but emerge in synthesized views. The quality of synthesized images is mainly contaminated by distortions in these disoccluded regions. Unfortunately, traditional image quality metrics are not effective for these synthesized images because they are sensitive to geometric distortions. To solve the problem, this paper proposes an objective quality evaluation method for 3D Synthesized images via Disoccluded Region Discovery (SDRD). A self-adaptive scale transform model is first adopted to preprocess the images on account of the impacts of view distance. Then disoccluded regions are detected by comparing the absolute difference between the preprocessed synthesized image and the warped image of preprocessed reference image. Furthermore, the disoccluded regions are weighted by a weighting function proposed to account for the varying sensitivities of human eyes to the size of disoccluded regions. Experiments conducted on IRCCyN/IVC DIBR image database demonstrate that the proposed SDRD method remarkably outperforms traditional 2D and existing DIBR-related quality metrics.

Index Terms— View synthesis, Image quality assessment, DIBR, SIFT flow, Disoccluded regions.

1. INTRODUCTION

The wide application of free-viewpoint video arouses the emergency of Depth-Image-Based-Rendering (DIBR) techniques, which aim to generate new viewpoints from multiple views. However, in the rendering process, these DIBR algorithms introduce distortions that mainly distribute at disoccluded regions in the synthesized images [1]. Unfortunately, traditional 2D image quality metrics are limited in the quality evaluation of synthesized images because of the geometric distortions.

So far, only a few works have been done for the quality assessment of synthesized images. Conze *et al*. [2] proposed to improve conventional 2D image quality methods by incorporating the maps of texture, gradient orientation and contrast as the weights. In [3], a full-reference synthesized image quality metric was presented. The synthesized and original views were first registered by blocks. Then the Kolmogorov-Smirnov distance between the blocks in both views was computed as the distortion of the block. Finally, the overall quality score was obtained by computing the distortions in all blocks. The Morphological Wavelet Peak Signal-to-Noise Ratio (MW-PSNR) metric proposed by Dragana *et al*. [4] first conducted the morphological bandpass wavelet decomposition. Then the Mean Squared Errors (MSE) at all levels were computed based on squared error maps. Finally, MW-PSNR is computed by pooling MSE. In [5], the same authors conducted a similar work, namely Morphological Pyramids Peak Signal-to-Noise Ratio (MP-PSNR) to improve MW-PSNR. However, both MP-PSNR and MW-PSNR are sensitive to geometric distortions, making them ineffective in assessing the quality of synthesized images. [6] proposed to measure the binocular asymmetry between the left and right images to evaluate the synthesized 3D images. While the performances of these methods are improved compared with conventional 2D metrics, they are far from being ideal.

In this paper, an objective quality assessment method for 3D Synthesized images via Disoccluded Region Discovery (SDRD) is presented. Considering the impacts of view distance, a self-adaptive scale transform [7] is first used to preprocess the synthesized and original images. The disoccluded regions are detected because the synthesized images are mostly distorted by the disoccluded areas. In the extraction process, we first warp the original image onto the synthesized image to compensate for the geometric displacement distortions. Then the absolute error map between the preprocessed synthesized and warped images is obtained. Furthermore, the error map is weighted to account for the varying sensitivities of human eyes to the size of distortion regions. Final-

This work is supported by the National Natural Science Foundation of China (61379143) and the Fundamental Research Funds for the Central Universities (2015XKMS032, 2015QNA66).

(a) (b)

Fig. 1. An example of disoccluded areas. (a) Original image; (b) Synthesized image with distortions marked by green boxes.

ly, the mean value of the weighted error map is computed as the overall quality score. Experiments are conducted on IRCCyN/IVC DIBR image database [1]. The results demonstrate that our work outperforms conventional 2D and existing DIBR-orientated image quality metrics.

2. PROPOSED SYNTHESIZED IMAGE QUALITY METRIC

The quality of synthesized images is mainly determined by the distortions in disoccluded regions. Fig. 1 shows an example to illustrate the impacts of disoccluded regions. It can be observed that compared with the original image (a), the synthesized image (b) is distorted by the obvious artifacts in disoccluded regions, which are marked by green boxes.

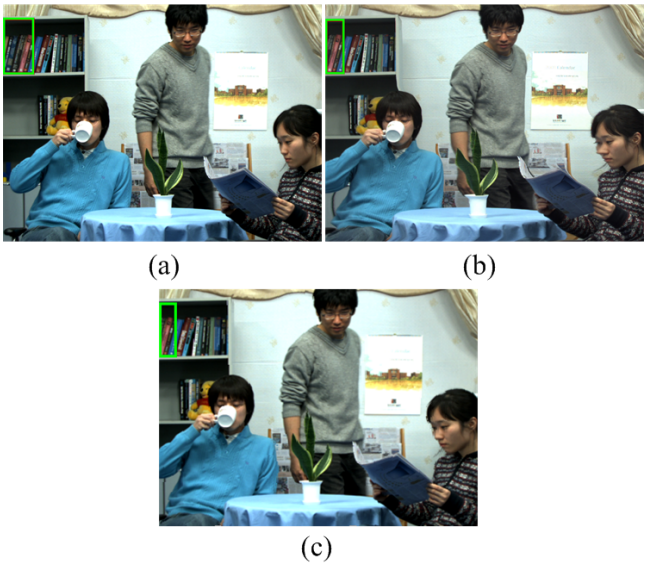
The proposed metric consists of three modules: a self-adaptive scale transform (SAST) module, a disoccluded region extraction module and a weighting module. The SAST module and the weighting module are used to account for the influence of view distance to image quality and the human visual sensitivities to the distortions with large size respectively.

2.1. Self-Adaptive Scale Transform

As mentioned above, the subjective image quality is influenced by view distance. In order to adapt to this characteristic, the input original and synthesized images $\mathbf{I}_1$, $\mathbf{I}_2$ are preprocessed into $\mathbf{I}'_1$, $\mathbf{I}'_2$ using a Self-Adaptive Scale Transform (SAST) model [7]. More advanced relevant technologies will be taken into consideration later [8]-[9].

SAST model is designed to transform the image to an optimal scale according to the image size and view distance. In this scale transformation process, the scaling coefficient z is computed by the following formula:

$$z = \sqrt{\frac{\mathcal{A}}{\widetilde{\mathcal{A}}}} = \sqrt{\frac{h * w}{\widetilde{h} * \widetilde{w}}} = \sqrt{\frac{1}{4\tan(\frac{\theta_h}{2}) * \tan(\frac{\theta_w}{2})} * (\frac{h}{d})^2 * r}, \quad (1)$$



(a) (b) (c)

Fig. 2. An example of warped results. (a) Original image; (b) Synthesized image with obvious geometric displacement from its original image (a); (c) The warped result of (a).

where d is the view distance; A and $\widetilde{A}$ represent the image size and visual scope respectively; h and $\widetilde{h}$ stand for objective and subjective view heights of the test image; similarly, w, $\widetilde{w}$ are widths of the objective and visual images; θ_h and θ_w are visual angles of both horizontal and vertical orientations; r is the ratio between the image height and width, i.e. $r = \frac{w}{h}$.

2.2. Disoccluded Region Detection

In general, disoccluded regions can be detected using the depth maps of the reference view and synthesized view. However, this method is invalid in some cases like the DIBR algorithm 'A1' in [1], which cropped the disoccluded regions and then interpolated it to its original size. Therefore, general disoccluded region detection method suitable for all the DIBR algorithms is required. In this paper, considering the geometric displacement between synthesized and original images, the SIFT flow method [10] is first adopted to warp the original image onto the synthesized image through estimating the dense correspondence between them.

SIFT images of $\mathbf{I}'_1$ and $\mathbf{I}'_2$ are first computed and the SIFT flow map $\mathbf{F}$ between SIFT images is generated. Then $\mathbf{I}'_1$ is warped onto $\mathbf{I}'_2$ using $\mathbf{F}$ and the warped image is denoted by $\mathbf{I}'_{1w}$. In Fig. 2, (a) is an original image after preprocessing and (b) is a preprocessed synthesized image with geometric displacement. We warp (a) onto (b) using the SIFT flow method and the warped image is shown in (c). From (c), we can see the effectiveness of the warp process used.

Towards obtaining the disoccluded regions, we compute the absolute error map between $\mathbf{I}'_{1w}$ and $\mathbf{I}'_2$, which is denoted by $\mathbf{E}$. Generally, the error map not only shows notable distor-

tions in disoccluded areas but also includes some weak distortions caused by geometric distortions, which are too faint to draw human attention. Therefore, we use an empirical threshold T (equals 50 in this paper) to partition the disoccluded regions $\mathbf{D}$ from the error map:

$$\mathbf{D}(x,y) = \begin{cases} 0, & if\ \mathbf{D}(x,y) < T, \\ \mathbf{E}(x,y), & otherwise, \end{cases} \tag{2}$$

where $1 \le x \le M$, $1 \le y \le N$, $M \times N$ denotes the size of image $\mathbf{I}_1'$.

Fig. 3 illustrates the detection process of disoccluded regions. It is observed from (d) that the disoccluded regions are correctly detected, which demonstrates the feasibility and validity of the proposed detection method.

2.3. Weighting Method

Since large disoccluded regions have more impacts on the perceived quality, a weighting function for disoccluded regions is exploited. We first divide $\mathbf{D}$ into overlapping blocks of size $n \times n$ ($n = 5$) centered on each pixel. Then the number of pixels belonging to the distorted part in each block is computed and suppose the number in block B, which is centered on (x, y), is n_1. The ratio of pixels belonging to the distorted regions in block B is denoted by $\mathbf{R}(x, y)$:

$$\mathbf{R}(x,y) = \frac{n_1}{n \times n}. \tag{3}$$

2.4. Pooling

The ratio map $\mathbf{R}$ is treated as the weighting map of $\mathbf{D}$, so the final quality score Q is calculated as:

$$Q = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\mathbf{R}(x,y)\mathbf{D}(x,y)}{MN}. \tag{4}$$

Note that higher Q value indicates the poorer image quality.

3. EXPERIMENTAL RESULTS

3.1. Experimental Settings

The performance of the proposed method is evaluated in IRCCyN/IVC DIBR image database [1]. It consists of 12 original images and their corresponding 84 synthesized views, which are generated using seven DIBR approaches. A discrete rating scale from one to five is adopted in its subjective experiment and the subjective scores of these images in the database are provided in the form of Mean Opinion Score (MOS). However, the methodology of subjective experiment is the Absolute Category Rating-Hidden Reference (ACR-HR) method, which means no reference images are included, so that it is unreasonable to evaluate the objective full-reference metrics using the MOS values directly. Therefore, the Difference Mean Opinion Score (DMOS) is calculated from MOS values of the hidden Source Reference Channel (SRC) and the Processed Video Sequence (PVS) [11]:

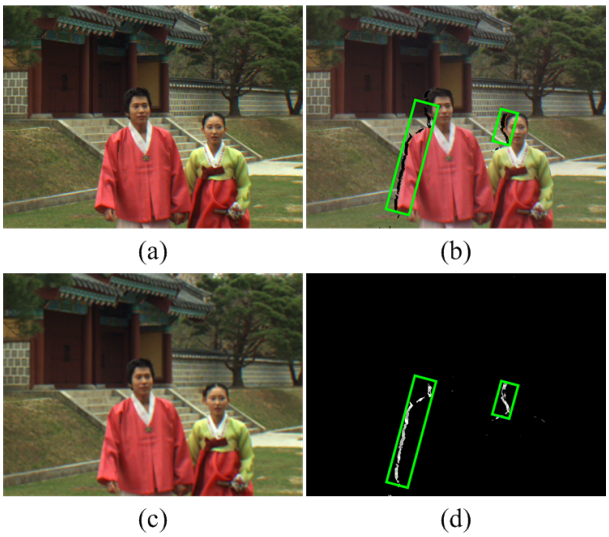


Fig. 3. An example to show the detection process of disoccluded regions. (a) An original image; (b) A synthesized image with disoccluded regions marked by green boxes; (c) Warped image of (a); (d) Disoccluded regions detected.

$$DMOS = MOS(PVS) - MOS(SRC) + 5, \tag{5}$$

where $MOS(PVS)$ and $MOS(SRC)$ are MOS values of the synthesized and corresponding original images respectively. In this formula, higher DMOS value represents better quality of synthesized image.

For performance evaluation, three recognized criterions, namely Pearson Linear Correlation Coefficient (PLCC), Root Mean Square Error (RMSE) and Spearman Rank order Correlation Coefficient (SRCC) are adopted. PLCC and RMSE are used to measure the prediction accuracy and SRCC is adopted to evaluate the prediction monotonicity. Higher values of PLCC and SRCC and lower value of RMSE indicate better quality. They are computed following a four-parameter nonlinear fitting:

$$f(x) = \frac{\lambda_1 - \lambda_2}{1 + e^{(x-\lambda_3)/\lambda_4}} + \lambda_2, \tag{6}$$

where $\lambda_i, i = 1, 2, 3, 4$ are the parameters to be fitted.

3.2. Performance Evaluation

Table 1 shows the experimental results of eleven 2D image quality methods and four existing DIBR-targeted methods in IRCCyN/IVC DIBR database. The 2D image quality metrics

Table 1. Performances of conventional 2D image quality metrics and existing DIBR-based methods in DIBR database.

Metrics	Types	PLCC	SRCC	RMSE
PSNR	2D	0.4279	0.4610	0.6018
SSIM [12]	2D	0.3703	0.3069	0.6185
MS-SSIM [13]	2D	0.4221	0.3113	0.6036
IW-SSIM [14]	2D	0.4809	0.3960	0.5838
VSNR [15]	2D	0.4012	0.4293	0.6614
MAD [16]	2D	0.4628	0.5332	0.5902
FSIM [17]	2D	0.4671	0.3286	0.5887
VSI [18]	2D	0.4370	0.3159	0.5989
GSM [19]	2D	0.4746	0.3402	0.6005
GMSD [20]	2D	0.4890	0.3338	0.5951
ADD-SSIM [21]	2D	0.6470	0.5611	0.5077
VSQA [2]	DIBR	0.5923	0.6010	0.5413
3Dswim [3]	DIBR	0.6623	0.6158	0.4988
MW-PSNR [4]	DIBR	0.5951	0.6246	0.5351
MP-PSNR [5]	DIBR	0.6148	0.6274	0.5251
Proposed SDRD	DIBR	**0.8104**	**0.7610**	**0.3901**

include PSNR, SSIM [12], MS-SSIM [13], IW-SSIM [14], VSNR [15], MAD [16], FSIM [17], VSI [18], GSM [19], GMSD [20], ADD-SSIM [21], while four DIBR metrics are VSQA [2], 3Dswim [3], MW-PSNR [4], MP-PSNR [5].

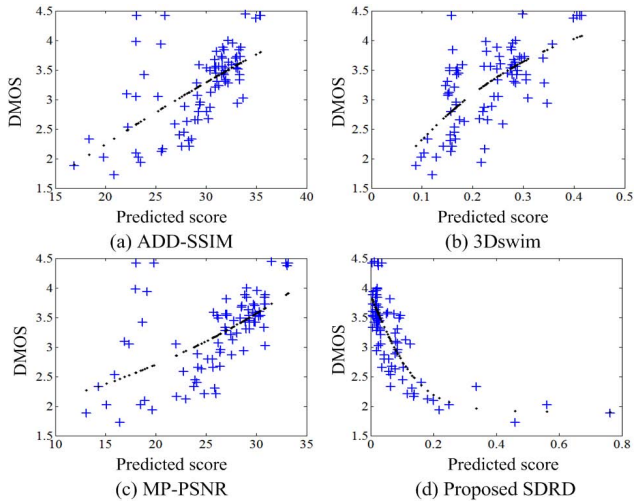


Fig. 4. Scatter plots of subjective scores versus the predicted scores of ADD-SSIM [21], 3Dswim [3], MP-PSNR [5] and the proposed method in DIBR database.

It is observed from Table 1 that conventional 2D image quality metrics are not effective in evaluating the quality of synthesized images. The best result of PLCC is below 0.65 and the maximum SRCC is only 0.5611. Existing DIBR algorithms [2-5] produce better results and the highest values

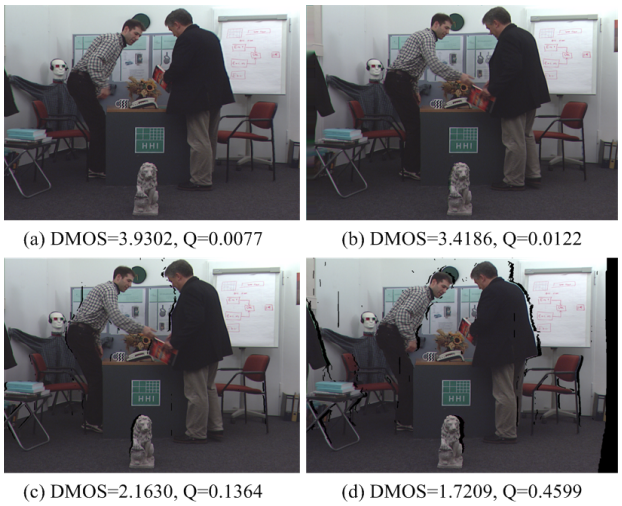


Fig. 5. Subjective and predicted scores of four synthesized images with different quality.

of PLCC and SRCC are 0.6623 and 0.6274 respetively. As contrast, the proposed method achieves the best performance, and PLCC and SRCC are 0.8104, 0.7610 respectively.

In order to further show the superiority of our work, the scatter plots between the predicted scores of top four methods in Table 1 and the subjective scores (DMOS) on the DIBR database are shown in Fig. 4. In these scatter plots, the points generated by a good metric are expected to distribute intensively around the fitting curve. Fig. 4 shows the sample points obtained from our metric all gather more closely around the fitting curve than other methods, which demonstrates the predicted scores of our method are more consistent with the subjective scores.

Fig. 5 shows four synthesized images of different distortion levels with their subjective and predicted scores. From (a)-(d), image quality is getting worse with declining DMOS. Correspondingly, Q value is becoming higher, which is consistent with subjective ratings.

4. CONCLUSION

In this paper, we have proposed an objective image quality assessment metric for synthesized images. The presented method consists of three parts, which are self-adaptive scale transform module, disoccluded region extraction module and a weighting module. The first part is employed on account of the influence of view distance. Then we propose to extract the disoccluded regions, which contaminate synthesized images mostly. Meantime, a weighting function is proposed to account for the human visual sensitivity to the distortions with large size. The experimental results conducted on IRCCyN/IVC DIBR image database have demonstrated the effectiveness and advantages of the proposed method.

5. REFERENCES

[1] E. Bosc, R. Pépion and P. L. Callet, "Towards a new quality metric for 3-D synthesized view assessment," *IEEE J. Select. Top. Signal Process.*, vol. 5, no. 7, pp. 1332-1343, Sep. 2011.

[2] P. H. Conze, P. Robert and L. Morin, "Objective view synthesis quality assessment," *Electron. Imag. Int. Society for Optics and Photonics*, vol. 8288, pp. 8288-8256, Feb. 2012.

[3] F. Battisti, E. Bosc, M. Carli and P. L. Callet, "Objective image quality assessment of 3D synthesized views," *Signal Process: Image Commun.*, vol. 30, pp. 78-88, Jan. 2015.

[4] D. S. Stankovic, D. Kukolj and P. L. Callet, "DIBR synthesized image quality assessment based on morphological wavelets," *Pro. IEEE Int. Workshop on Quality of Multimedia Experience*, pp. 1-6, Jan. 2015.

[5] D. S. Stankovic, D. Kukolj and P. L. Callet, "DIBR synthesized image quality assessment based on morphological pyramids," *The True Vision-Capture, Transmission and Display of 3D Video*, pp. 1-4, Oct. 2015.

[6] Y. J. Jung, H. G. Kim and Y. M. Ron, "Critical binocular asymmetry measure for perceptual quality assessment of synthesized stereo 3D images in view synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, DOI: 10.1109/TCSVT.2015.2430632.

[7] K. Gu, G. T. Zhai, X. K. Yang and W. J. Zhang, "Self-adaptive scale transform for IQA metric," *Proc. IEEE Conf. Int. Symp. Circuits Syst.*, pp. 2365-2368, May 2013.

[8] K. Gu, G. T. Zhai, M. Liu, Q. Xu, X. K. Yang, and W. J. Zhang, "Adaptive high-frequency clipping for improved image quality assessment," *Proc. IEEE Conf. Vis. Commun. Image Process.*, pp. 1-6, Nov. 2013.

[9] K. Gu, M. Liu, G. T. Zhai, X. K. Yang, and W. J. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Trans. Broadcasting*, vol. 61, no. 3, pp. 520-531, Sep. 2015.

[10] C. Liu, J. Yuen, A. Torralba, J. Sivic and W. T. Freeman, "SIFT flow: dense correspondence across different scenes," *Proc. European Conf. Comput. Vis.*, vol. 5304, pp. 28-42, Oct. 2008.

[11] G. Cermak, L. Thorpe and M. Pinson, "Test plan for evaluation of video quality models for use with high definition TV content," *Video Quality Experts Group*, 2009.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[13] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment," *in Pro. Asilomar Conf. Signals, Systems, and Computers*, vol. 2, pp. 1398-1402, Nov. 2003.

[14] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185-1198, May 2011.

[15] D. Chandler and S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284-2298, Sep. 2007.

[16] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full reference image quality assessment and the role of strategy," *J. Electronic Imaging*, vol. 19, no. 1, pp. 001006:1-21, Mar. 2010.

[17] L. Zhang, L. Zhang, X. Q. Mou and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378-2386, Aug. 2011.

[18] L. Zhang, Y. Shen and H. Li, "VSI: a visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270-4280, Oct. 2014.

[19] A. M. Liu, W. S. Lin and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500-1512, Apr. 2012.

[20] W. F. Xue, L. Zhang, X. Q. Mou and A. C. Bovik, "Gradient magnitude similarity deviation: a highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684-695, Feb. 2014.

[21] K. Gu, S. Q. Wang, G. T. Zhai, W. S. Lin, X. K. Yang, and W. J. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Trans. Broadcasting*, DOI: 10.1109/TBC.2015.2511624.