

SC-IQA: Shift compensation based image quality assessment for DIBR-synthesized views

Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges

► **To cite this version:**

Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges. SC-IQA: Shift compensation based image quality assessment for DIBR-synthesized views. VCIP 2018, Dec 2018, Taichung, Taiwan. 10.1109/vcip.2018.8698654 . hal-01899637

HAL Id: hal-01899637

<https://hal.archives-ouvertes.fr/hal-01899637>

Submitted on 19 Oct 2018

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.

SC-IQA: Shift compensation based image quality assessment for DIBR-synthesized views

Shishun Tian ^{*1}, Lu Zhang ^{*2}, Luce Morin ^{*3} Olivier Déforges ^{*4}

^{*} IETR UMR CNRS 6164, National Institute of Applied Sciences of Rennes, France

20, Avenue des Buttes de coësmes, 35708 Rennes Cedex 7, France

^{1,3,4} `firstname.lastname@insa-rennes.fr`

² `lu.ge@insa-rennes.fr`

Abstract—Depth-image-based-rendering (DIBR) has been used to generate the virtual views for Multi-view videos and Free-viewpoint videos. However, the quality assessment of DIBR-synthesized views is very challenging owing to the new types of distortions induced by inaccurate depth maps, dis-occlusions and image inpainting methods. There exist a large number of object shifts and geometric distortions in the synthesized view which the traditional 2D quality metrics may fail to assess. In this paper, we propose a shift compensation based image quality assessment metric (SC-IQA) for DIBR-synthesized views. Firstly, the global geometric shift is compensated roughly by an SURF + RANSAC homography approach. Then, a multi-resolution block matching method, which performs a more accurate matching, is used to precisely compensate the shift and penalize the local geometric distortion as well. In addition, a visual saliency map is also used as a weighting function. To calculate the final overall quality scores, only the worst blocks are utilized since the biggest distortions have the most effects on the overall perceptual quality. The results show that the proposed metric significantly outperforms the state-of-the-art synthesized view dedicated metrics and the conventional 2D IQA metrics.

Index Terms—Synthesized view, Quality assessment, Shift compensation, Affine transform, Block matching

I. INTRODUCTION

Enabling users to view a 3D scene from different angles, Free-viewpoint TV (FTV) and Multiview TV have gained great public interest and curiosity in the past decade. Multiview-Video-Plus-Depth (MVD) is the main representation method for 3D applications. It consists of limited number of texture views from different viewpoints and their associated depth maps, which are used to synthesize the virtual views at other viewpoints based on Depth-Image-Based-Rendering (DIBR) [1]. However, view synthesis can also induce some new types of distortion due to depth errors, dis-occlusions and inpainting methods. The distortions caused by 2D image compression, such as blur, blockiness and ringing, are often scattered over the whole image while the DIBR-synthesized distortions are mostly local. Besides, inaccurate depth maps and imperfect image inpainting methods can also introduce object shifting and geometric distortions in the synthesized views. For this reason, the conventional 2D quality metrics do not work well on DIBR-synthesized artifacts. Thus, quality

metrics suitable for synthesized views are urgently needed. Such an effective quality metric could be used as a feedback to optimize not only the view synthesis but also the compression.

In order to solve this problem, several full-reference (FR) methods have been proposed to evaluate the quality of DIBR-synthesized views in the past few years.

- 1) VSQA: View Synthesis Quality Assessment [2] was proposed aiming to characterize the image complexity by applying three weighting maps on the SSIM distortion map. The gain of PLCC approached 17.8% over original SSIM in correlation with subjective judgment.
- 2) 3DSwIM: 3D Synthesized view Image Quality Metric [3] is based on a comparison of statistical features of wavelet sub-bands of the original and DIBR-synthesized images. A registration process and a skin detection step are used to make sure that the best matching blocks are always compared, and that the most sensitive distorted blocks are penalized.
- 3) MW-PSNR, MP-PSNR: Morphological Wavelet Peak Signal-to-Noise Ratio measure [4] and Morphological Pyramid Peak Signal-to-Noise Ratio [5] were proposed by Sandic-Stankovic et al. to handle the geometric distortions in different resolution levels. In addition, their reduced versions, the reduced MW-PSNR (MWPSNRr) and reduced MP-PSNR (MP-PSNRr) in [6], were also proposed to improve the performance.
- 4) CT-IQA: a variable-length context tree based image quality assessment [7] proposed by Ling et al. aims to quantify the overall structure dissimilarity and dissimilarities in various contour characteristics.
- 5) EM-IQA: In [8], Ling et al. also proposed an elastic metric based image quality assessment metric by quantifying the deformation of curves in the local distortion regions.

According to recent research [9], Human Visual System (HVS) is more sensitive to local artifacts compared to the global object shift. However, the global shift in DIBR-synthesized views can be easily penalized by most pixel-wise quality metrics, eg. PSNR, SSIM. In this paper, we propose an FR shift compensation based image quality assessment metric (SC-IQA) for DIBR-synthesized views. Besides, we use a visual saliency map to weight the final quality map. This paper

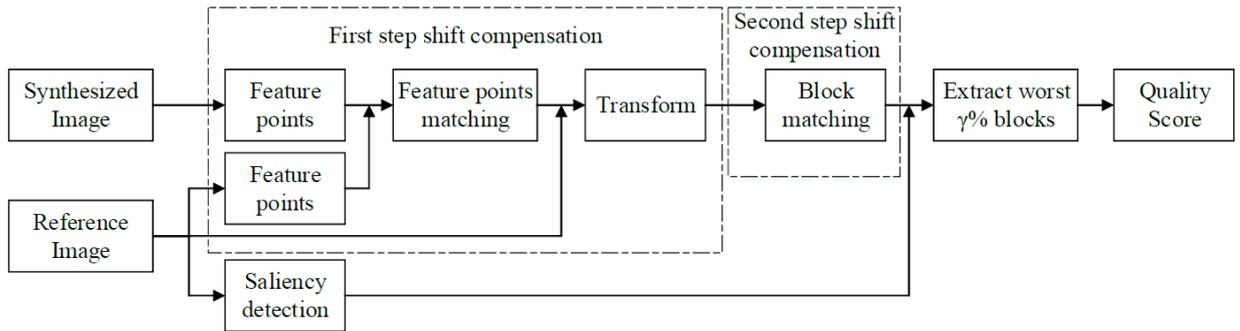


Fig. 1. Block scheme of the proposed method

is organized as follows: the proposed method is introduced in Section II, the experimental results on the IRCCyN/IVC DIBR database [10] are presented and discussed in Section III and finally the conclusions are drawn in Section IV.

II. PROPOSED METHOD

In this section, the proposed FR metric is introduced in detail. We assume that the reference views, which is captured at the same viewpoints as the synthesized views, are available; and the FR metrics compute the distortions between the synthesized and the reference views. As mentioned above, the global shift is acceptable to human observer while it can be easily penalized by pixel-wise metrics. Thus, the proposed method concentrates on compensating the global shift between the DIBR-synthesized views and the reference views.

The block diagram is shown in Fig. 1, the proposed shift compensation method consists of two parts: firstly, the global geometric shift is compensated approximately by feature points matching and transform is used to compensate the shift roughly; then, we use block matching to compensate it precisely. Besides, a saliency map is also used as a weighting function to improve the performance. The final overall quality scores are obtain by measuring the γ % worst matched blocks since human observers are more sensitive to poor quality regions rather than the good ones.

A. Feature points matching and transform

Fig. 2 (a) gives an example of the SSIM map between the synthesized image and the reference image in the adopted database [10], it could be observed that there exists great global shift between the synthesized image and the reference image. In this part, the global geometric shift is compensated roughly by SURF [11] + RANSAC [12] homography approach.

Firstly, SURF feature points in the reference and synthesized images are detected and matched. Then, to be robust, the RANSAC algorithm is used to refine the matching and estimate the homography matrix H . After that, the pixels of the synthesized image are warped to the corresponding positions in reference image by H . The SSIM map before transform, the matched feature point pairs and the SSIM map after transform are shown in Fig. 2.

We can observe that the global shift between the synthesized and the reference images has been roughly compensated. In the next step, we use block matching to compensate the shift precisely.

B. Block matching

In this part, a multi-resolution block matching algorithm is used to precisely compensate the shift and also to detect the large geometric distortions. In the first step, we use a large block $N1 \times N1$ ($N1 = 64$) for primary matching; then we use a small block $N2 \times N2$ ($N2 = 8$) for final matching. The matching process can be described by the following steps:

- 1) Divide the synthesized view into a regular grid of $N1 \times N1$ blocks;
- 2) For each $N1 \times N1$ block, search for the best matching block in the reference view. The best matching block is the one showing the largest following similarity criterion:

$$sim(s, r) = \frac{cov(s, r) + \epsilon}{var(s) + var(r) + \epsilon} \quad (1)$$

where s, r denote the blocks in the synthesized image and the reference image; the operation cov and var denote the co-variance and variance respectively; ϵ is a constant value to stabilize the division with weak denominator.

- 3) Each $N1 \times N1$ block is divided into smaller $N2 \times N2$ blocks and the process is repeated with a smaller search window.

Since the shift only occurs in the horizontal direction, we only search the blocks in this direction for matching. We assume the biggest shift in the synthesized image to be 30, the search windows of $N1$ and $N2$ are restricted to 30 and 5 respectively.

The goal of this multi-resolution block matching algorithm is to compensate the global shift and penalize the local geometric distortions. Now, if we directly use $N2 \times N2$ block for matching, and set the search window to 30 (the biggest shift range in the synthesized image), the computational complexity will be much higher. Besides, as shown in Fig. 3, there exists great geometric distortion in the red block ($N1 \times N1$) in Fig. 3 (a) compared to its matched block in the reference



(a) SSIM map before transform (b) optimized matched feature point pairs (c) SSIM map after transform

Fig. 2. Example of feature points matching and transform

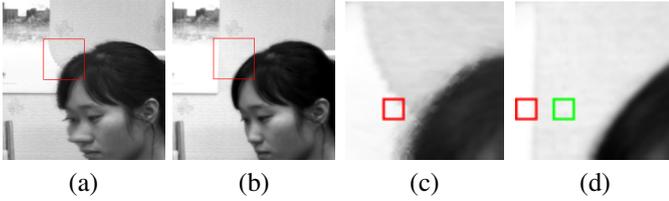


Fig. 3. Block matching: (a), (b) are the patches in the synthesized and the reference image; (c) block in the synthesized image; (d) matched block in the reference image

image (the red block in Fig. 3 (b)). If we directly use 8×8 block for matching and set the searching window to 30, the best matched block for the block in Fig. 3 (c) is the red block in Fig. 3 (d). There exists little difference between these two red blocks, so the geometric distortion will not be penalized. On the contrary, if we use the proposed multi-resolution block matching method, the matched block is the green one, this geometric distortion will be surely penalized. The multi-resolution approach is thus more robust.

C. Saliency weighting

In addition, a saliency detection [13] is also used as a weighting map to improve the performance of the proposed metric. The distortion of each $N_2 \times N_2$ block is measured by averaging the weighted mean square errors between the blocks of the synthesized and the reference images, as shown in:

$$MSE_B = \frac{\sum_{(i,j) \in B} (syn(i,j) - ref(i,j))^2 \times Salmap(i,j)}{\sum_{(i,j) \in B} Salmap(i,j)} \quad (2)$$

where B means the matched $N_2 \times N_2$ blocks; (i, j) denotes the pixel in the block; syn and ref represent the blocks in the synthesized image and reference image respectively; $Salmap$ represents the saliency map in this block.

D. Quality pooling

Since the humans tend to perceive poor regions in an image with more severity than the good ones [9], [14], we only use the blocks with the worst quality to calculate the final quality

as shown in Eq. 3.

$$MSE_W = \frac{1}{N_W} \sum_{i \in W} MSE_B(i) \quad (3)$$

where W represents the set of the worst 1% blocks in the image, N_W is the number of items in the set W .

The final quality scores are computed as the following equation:

$$Score_{SC-IQA} = 10 \times \log_{10}(255 \times 255 / MSE_W) \quad (4)$$

where a higher quality score indicates a better quality.

III. EXPERIMENTAL RESULTS

This section describes and discusses the experiments of the proposed SC-IQA metric. The performance of SC-IQA is tested on the IRCCyN/IVC DIBR database [10], which consists of 84 synthesized views generated by seven different DIBR view synthesis algorithms and their associated 12 reference views along with the subjective scores - mean opinion score (MOS).

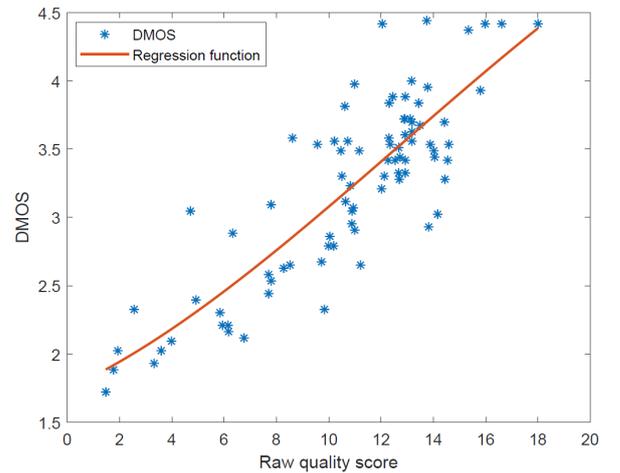


Fig. 4. Scatter plot of SC-IQA quality score versus DMOS

The performance of objective quality metrics can be represented by their correlations with subjective judgements. In this paper, the following 3 widely employed criteria are used:

Pearson Linear Correlation Coefficients (PLCC), Spearman Rank Order Correlation Coefficients (SROCC) and Root-Mean-Square-Error (RMSE). Before calculating these 3 coefficients, the objective scores need to be regressed to remove the nonlinearities due to the subjective rating process, as recommended by Video Quality Expert Group (VQEG) Phase I FR-TV [15]. The scatter plot of the predicted scores versus the subjective scores and the regression function are shown in Fig. 4.

The comparison results of the proposed metric (with $\gamma = 1$), the state-of-the-art DIBR dedicated quality metrics and the commonly used 2D quality metrics are concluded in Table I. It shows that the proposed metric significantly outperforms the state-of-the-art DIBR dedicated quality metrics: MW-PSNR, MP-PSNR, MP-PSNRr, MW-PSNRr, 3DSwIM, VSQA, CT-IQA and EM-IQA, and the commonly utilized 2D image quality metrics: PSNR, SSIM.

TABLE I
PERFORMANCE COMPARISON BETWEEN THE PROPOSED METRIC AND THE STATE-OF-THE-ART METRICS (WITH $\gamma = 1$)

| Metric | PLCC | RMSE | SROCC |
|---------------|---------------|---------------|---------------|
| PSNR | 0.4557 | 0.5927 | 0.4417 |
| SSIM | 0.4348 | 0.5996 | 0.4004 |
| 3DSwIMr | 0.6864 | 0.4842 | 0.6125 |
| VSQA | 0.6122 | 0.5265 | 0.6032 |
| MW-PSNR | 0.6200 | 0.5224 | 0.5739 |
| MW-PSNRr | 0.6625 | 0.4987 | 0.6232 |
| MP-PSNR | 0.6729 | 0.4925 | 0.6272 |
| MP-PSNRr | 0.6954 | 0.4784 | 0.6606 |
| CT-IQA | 0.6809 | 0.4877 | — |
| EM-IQA | 0.7430 | 0.4455 | — |
| SC-IQA | 0.8496 | 0.3511 | 0.7640 |

In addition, the performance dependency of the proposed metric on ratio $\gamma\%$ is also discussed as shown in Fig. 5. Especially the ratio 0 is associated with the proposed metric without saliency map. According to Fig. 5, the use of the saliency map improves the proposed metric slightly; the performance of the proposed metric goes down as the ratio increases, however even the lowest scores are still superior to most of the state-of-the-art metrics in Table I. This shows the robustness of the proposed metric.

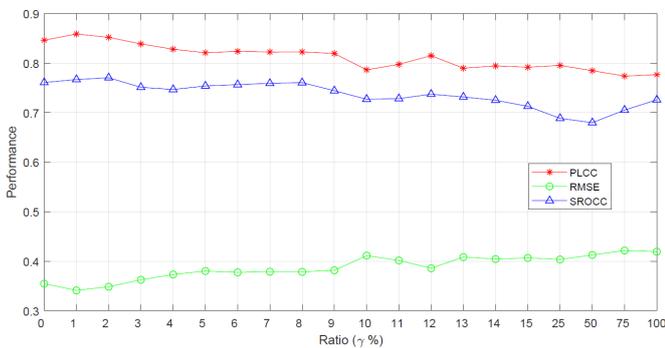


Fig. 5. Performance dependency of the proposed metric with the changing ratios ($\gamma\%$)

IV. CONCLUSION

In this paper, we proposed a novel full-reference shift compensation based image quality assessment metric for DIBR-synthesized views (SC-IQA). An SURF + RANSAC homography approach and a multi-resolution block matching are used to compensate the global shift and to penalize the geometric distortions as well. The experimental results show that it significantly outperforms the state-of-the-art 3D synthesized view dedicated metrics: 3DSwIM, MP-PSNR, MW-PSNR, CT-IQA, EM-IQA and the conventional 2D IQA metrics: PSNR, SSIM.

REFERENCES

- [1] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3d-tv," in *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291. International Society for Optics and Photonics, 2004, pp. 93–105.
- [2] P.-H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," in *Stereoscopic Displays and Applications XXIII*, vol. 8288. International Society for Optics and Photonics, 2012, p. 82881M.
- [3] 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.
- [4] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "Dibr synthesized image quality assessment based on morphological wavelets," in *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*. IEEE, 2015, pp. 1–6.
- [5] —, "Multi-scale synthesized view assessment based on morphological pyramids," *Journal of Electrical Engineering*, vol. 67, no. 1, pp. 3–11, 2016.
- [6] D. Sandić-Stanković, D. Kukolj, and P. 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.
- [7] S. Ling, P. Le Callet, and G. Cheung, "Quality assessment for synthesized view based on variable-length context tree," in *Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on*. IEEE, 2017, pp. 1–6.
- [8] S. Ling and P. Le Callet, "Image quality assessment for dibr synthesized views using elastic metric," in *Proceedings of the 2017 ACM on Multimedia Conference*. ACM, 2017, pp. 1157–1163.
- [9] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE journal of selected topics in signal processing*, vol. 3, no. 2, pp. 193–201, 2009.
- [10] 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.
- [11] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.
- [12] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Readings in computer vision*. Elsevier, 1987, pp. 726–740.
- [13] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 2083–2090.
- [14] X. Liu, Y. Zhang, S. Hu, S. Kwong, C.-C. J. Kuo, and Q. Peng, "Subjective and objective video quality assessment of 3D synthesized views with texture/depth compression distortion," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4847–4861, 2015.
- [15] 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.