NIQSV: A NO REFERENCE IMAGE QUALITY ASSESSMENT METRIC FOR 3D SYNTHESIZED VIEWS

Shishun TIAN, Lu ZHANG, Luce MORIN, Olivier DEFORGES

IETR UMR CNRS 6164, INSA de Rennes
20, Avenue des Buttes de coesmes, 35708 Rennes Cedex 7, France

ABSTRACT

The popularity of 3D applications, such as Free View-point TV (FTV) and Multi-view Video plus Depth (MVD), induces a heavy requirement of synthesized views. However, the quality assessment of synthesized views is very challenging because the corresponding original views (reference views) are usually not available at both encoder and decoder sides. In this paper, we propose a new no-reference quality assessment model to evaluate the quality of 3D synthesized views, called NIQSV (No-reference Image Quality assessment of Synthesized Views). This metric is based on the hypothesis that a good quality image is composed of flat areas (objects) separated by sharp edges, and the quality estimation involves only a set of simple morphological operators. NIQSV integrates the distortions of all the components, and then uses an edge image to weight the final distortions since the distortions of synthesized views mainly happen around object edges. The experimental results show that the proposed metric outperforms traditional 2D metrics and ranks among the best of dedicated 3D synthesized and full reference metrics.

Index Terms— FTV, synthesized views, quality assessment, no-reference, morphology

1. INTRODUCTION

Providing the depth perception of a visual scene, 3D video applications have gained great public interest and curiosity in the past decade. They are known as 3D-TV [1] and Free-viewpoint TV [2]. Free-viewpoint TV (FTV) is able to allow the users to view a 3D scene by freely changing their viewpoints. However, even with a very large number of views, it still can not cover all the arbitrary viewpoints of particular scenes [3]. The need of rendering the additional virtual view thus arises.

Depth-image-based-rendering (DIBR) has been used to generate the virtual views for several 3D applications, but this process can also create some new kinds of distortions, which are very different from that of 2D images. For this reason, most of the 2D objective quality metrics may fail in assessing the quality of DIBR synthesized images. On the other hand, the use of subjective tests may be expensive, time consuming, cumbersome and practically not feasible in systems where real-time quality score of an image or video sequence is needed. Hence, objective metrics are urgently needed for assessing the quality of synthesized images.

Many efforts have been made towards the assessment of synthesized images, such as VSQA [4], 3DSwIM [5], MW-PSNR [6] and MP-PSNR [7]. In [4], Conze et al. proposed VSQA a full-reference (FR) objective quality assessment metric which aims to handle the areas where disparity estimation may fail: thin objects, objects borders, transparency, variations of illumination or color differences between left and right views. VSQA uses 3 visibility maps to characterize complexity in terms of textures, diversity of gradient orientations and presence of high contrast. It achieves a gain of 17.8% over SSIM in correlation with subjective measurements. In [5], 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 used as the distortions of synthesized views mainly occur in the horizontal direction. A registration process is used to make sure that the best matching blocks are always compared, and a skin detection step is included to penalize distorted blocks containing ”skin-pixels” based on the assumption that the human observer is most sensitive to impairments affecting human subjects. 3DSwIM outperforms the conventional 2D metrics and existing DIBR image related metrics. Sandic-Stankovic et al. proposed a full-reference metric MP-PSNR [7] based on multi-scaled pyramid decomposition using morphological filters. The non-linear morphological filters maintain important geometric information such as edges across different resolution levels. Besides, the authors proposed a reduced MP-PSNR which only takes into consideration the mean squared errors between pyramids' images at higher scales. This metric is reported to achieve much higher correlation with human judgment compared to the state-of-art image quality assessment metrics. The same authors also proposed a MW-PSNR metric [6] relying on morphological decomposition, which achieved a higher correlation with human judgement.

The metrics introduced above are all full-reference (FR) metrics. However in some 3D applications, there is only a

E-mail address: shishun.tian@insa-rennes.fr.

limited number of viewpoints which are captured, coded and transmitted, so there is a large number of views which do not have reference views need to be synthesized. In other words, there is no ground truth (reference view) allowing a full comparison with the distorted view. As a result, no-reference quality assessment tools are in great need [8]. In order to handle this issue, we propose in this paper a new no-reference (NR) metric to estimate the quality of the synthesized images in the absence of the corresponding reference images.

Our metric is a morphological and edge based metric, the following section presents the principles of the proposed metric.

2. IMAGE AND VIEW SYNTHESIS MODEL

DIBR synthesized distortions mainly happen in the non-visible regions of previous view which become visible in the synthesized view. These artifacts can be produced by depth compression and the synthesis process. Some typical artifacts are introduced below:

- Object shiftings: object regions can be slightly shifted in the synthesized view owing to the depth preprocessings including low-passing filters or depth encoding methods to smooth the object borders.
- Incorrect rendering: rendering errors may occur in the complex textured areas which in-painting methods may fail to reconstruct.
- Blurry regions: some blurry regions may be produced by the in-painting method used to fill the disoccluded areas.
- Flickering: this is a temporal artifact which occurs in synthesized video due to the random errors happened in depth sequence.
- Crumbling: the object edge seems distorted in the synthesized view, this is mainly cased by the artifacts in depth data around strong discontinuities which appear like erosion.

we propse a new no-reference (NR) metric to assess the quality of 3D synthesized view, No-reference Image Quality assessment of Synthesized Views (NIQSV). It is based on the following image model: a good quality image is assumed to present sharp and regular object borders, smooth values inside the object and large discontinuities at the object borders. These “perfect” images are insensitive to morphological opening and closing operations, but some artifacts introduced above such as incorrect rendering, blurry regions and crumbling could be removed or detected by those morphological operations. This is the principle of our metric.

3. PROPOSED METRIC

In this paper, we propose a no-reference (NR) metric NIQSV, it quantifies the distortions in luminance, contrast and saturation using a set of morphological operations. Then, we integrated these three distortions of each component into one by a color weight factor k_c. Furthermore, an edge image is utilized to weight the final distortions since the distortions of synthesized views mainly happen around object edges. The block scheme is presented in Fig 1.

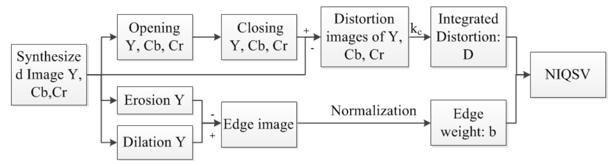


Fig. 1. Block scheme of NIQSV

The key strategy of NIQSV is a pair of opening and closing operations. The opening operation used on the synthesized image can help to remove some thin blurry regions, and the following closing operation with a relatively larger Structural Element(SE) can fill the holes in the disoccluded areas. It can be computed as follows,

$$D_X = |(I_X \circ SE) \bullet SE) - I_X|, X \in (Y, Cb, Cr) \quad (1)$$

where D_X presents the distortion of each color component, and I_X is the corresponding color component of the synthesized image.

In order to obtain the overall distortion, the distortions of all components are integrated as in Eq 2, where k_c presents the weight of color:

$$D = (1 - k_c) \cdot D_Y + \frac{k_c}{2} \cdot (D_{Cb} + D_{Cr}) \quad (2)$$

Since the artifacts mostly happen around the edges, the image edges must be taken into consideration. The edge image is obtained by morphological operators as described in Eq 3:

$$Edge = (I_Y \oplus SE) - (I_Y \ominus SE) \quad (3)$$

In order to get a weight factor, the Edge values are normalized to (0,1), factor a and b are computed as follows, where k_e presents the weight of edge:

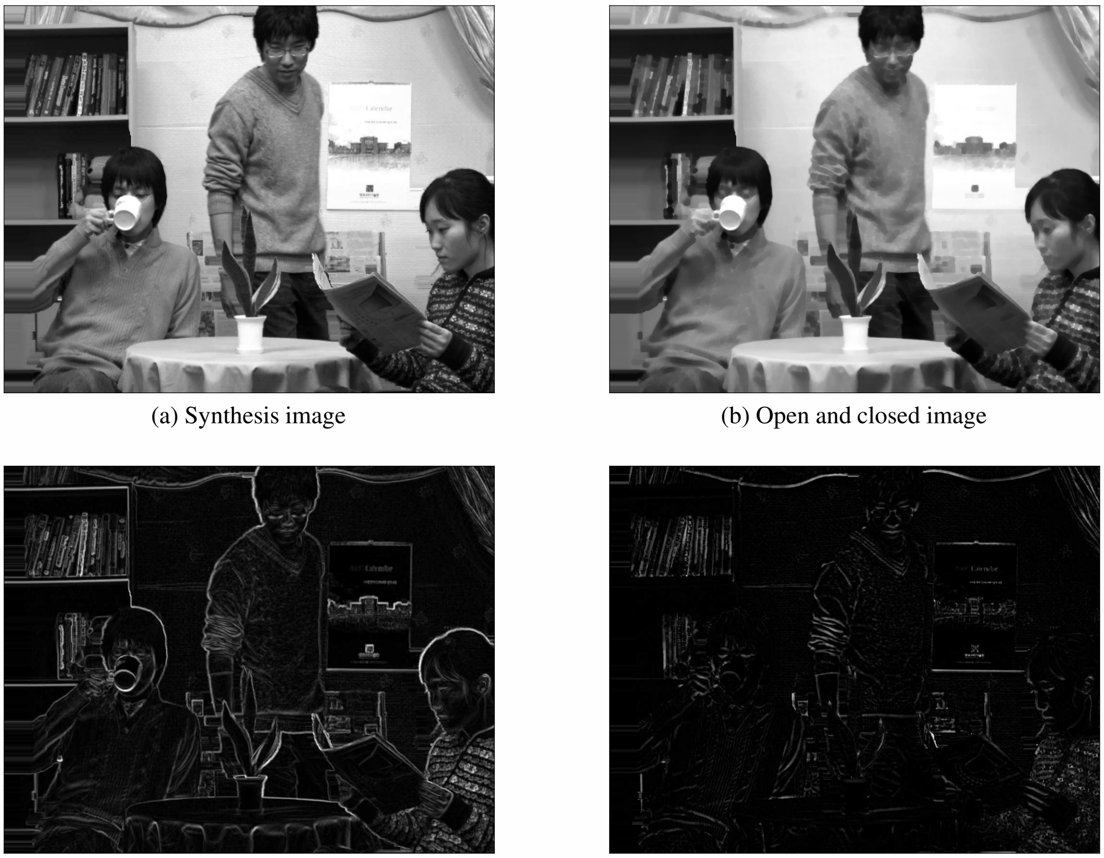
$$a = Edge/255, a \in (0, 1) \quad (4)$$

$$b = [(1 - k_e) + a \cdot k_e], k_e \in (0, 1) \quad (5)$$

Finally, the overall image quality score $NIQSV$ is computed as follows:

$$MSE' = \frac{\sum_{(i,j)\in I} b(i,j) \cdot D(i,j)^2}{\sum_{(i,j)\in I} b(i,j)} \quad (6)$$

$$NIQSV = 10 \cdot log_{10}(\frac{255 \times 255}{MSE'}) \quad (7)$$



(a) Synthesis image

(b) Open and closed image

(c) Normalized edge

(d) Integrated Distortion

Fig. 2. Newspaper sequence view 4 synthesized from 6 Synthesized with A4 [9]

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section describes the validation of the proposed NIQSV metric. The performance of this metric are evaluated using IRCCyN/IVC DIBR image database [10] which is introduced in [11]. It contains the frames from 3 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, there are four virtual views generated on the positions that obtained by the real cameras using seven DIBR synthesis algorithms A1-A7 as follows:

- A1 [12]: the depth map is filtered to remove the depth discontinuities, the borders are cropped and then the image is interpolated to reach its original size. This could lead to object shifting artifacts.
- A2: based on A1 except that the borders are in-painted as in [13] instead of cropping. There should be geometry distortions owing to the low-pass filter of depth map.
- A3: Tanimoto et al. [14] proposed a 3D view generation system which is adopted reference software for MPEG in 3D video group.
- A4: A4 performs a hole filling method aided by depth information. [9]
- A5: Ndjiki-Nya et al. [15] used a patch-based texture synthesis method to fill the missing part in the virtual view.
- A6: Koppel et al. [16] extended A5 by a background sprite which take the temporal information into consideration to improve the synthesis.
- A7: A7 refers to those sequences who keep the holes in virtual views unfilled.

According to our tests, the parameters who make NIQSV perform the best were used, while $k_c = 0.45$ and $k_e = 1$. Figure 2 shows some examples of processed images of Newspaper sequence view 4 synthesized from 6. Figure 3 gives two examples of the NIQSV values of synthesized images using A1 and A7.

The reliability of objective metrics can be evaluated by

(a) A1 NIQSV: 31.402 db (b) A7 NIQSV: 20.735 db

Fig. 3. NIQSV values of synthesized images by A1 and A7

their correlation between subjective test scores. Usually, we use Differential Mean Opinion Score (DMOS) as subjective score. In this paper, the consistency of objective metrics was calculated by using Pearson Linear Correlation Coefficients (PLCC), Spearman's Rank Order Correlation Coefficients (SROCC) and Root-Mean-Square-Error (RMSE). Before calculating PLCC, RMSE and SROCC, the objective scores needed to be fitted to the so-called predicted DMOS, which noted as $DMOS_p$. Video Quality Expert Group (VQEG) Phase I FR-TV [17] has recommended the following nonlinear mapping function for this fitting step:

$$DMOS_p = a \cdot scores^3 + b \cdot score^2 + c \cdot score + d \quad (8)$$

while $score$ is the score obtained by the objective metric and a, b, c, d are the parameters of this cubic function. They are obtained through regression to minimize the difference between $DMOS_p$ and $DMOS$. Figure 4 shows the comparison between $DMOS$ and the fitted scores $DMOS_p$.

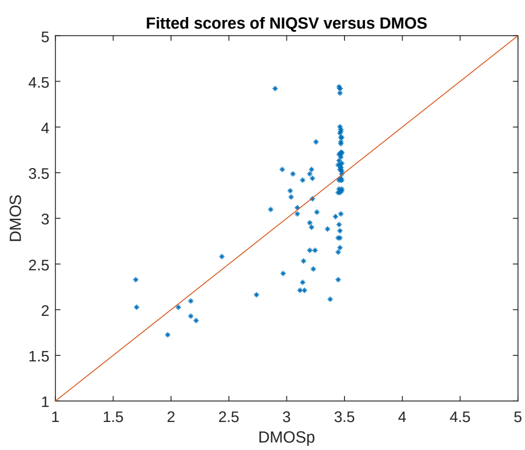


Fig. 4. Fitted scores vesus DMOS

Table 1 gives the obtained PLCC, RMSE, SROCC values and shows that NIQSV performs much better than PSNR and SSIM and achieves very closely to the other three full-reference metrics: 3DSwIM, MW-PSNR and MP-PSNR, the SROCC value is even a little better than 3DSwIM.

Table 1. PLCC, RMSE and SROCC between DMOS and objective metrics

Metric	PLCC	RMSE	SROCC
NIQSV (NR)	0.6346	0.5146	0.6167
3DSwIM (FR)	0.6864	0.4842	0.6125
MP-PSNR (FR)	0.6954	0.4784	0.6606
MW-PSNR (FR)	0.6737	0.4921	0.6493
PSNR (FR)	0.4117	0.6068	0.3514
SSIM (FR)	0.2665	0.6417	0.1832

Table 2. Ranking of synthesis algotithms according to DMOS and objective metrics

Metric	Ranking of synthesis algorithms						
DMOS	A1	A5	A4	A6	A2	A3	A7
NIQSV (NR)	A1	A4	A5	A2	A6	A3	A7
3DSwIM (FR)	A1	A4	A5	A6	A3	A2	A7
MW-PSNR (FR)	A4	A5	A6	A2	A3	A1	A7
MP-PSNR (FR)	A4	A5	A6	A3	A2	A1	A7
PSNR (FR)	A6	A5	A4	A3	A2	A7	A1
SSIM (FR)	A3	A4	A5	A6	A2	A7	A1

Moreover, in order to compare the performance of these image quality metrics, we also noted the rankings of the performance of synthesis algorithms according to DMOS and objective metrics in this Database. As shown in Table 2, the synthesis algorithm ranked in the first column holds the best performance compared to those located in the last column at the right side performs the worst. The first line offers the rankings according to DMOS scores which provides the ground truth, the following three lines give the rankings of the proposed metric and the other two DIBR synthesis image dedicated metrics, and the remaining lines present the performance of some 2D image quality assessment metrics. We can see that the proposed metric ranks very closely to DMOS scores except A5/A4 and A6/A2. Considering A5/A4 and A6/A2 are very closed in terms of DMOS, NIQSV ranks much better than dedicated 3D synthesized and full reference metrics.

5. CONCLUSION

In this paper, we proposed a new no-reference quality assessment metric for 3D synthesized views based on a set of morphological operations. The experimental results show that the proposed metric outperforms the traditional 2D metrics and approaches the results of 3D synthesized view aimed full reference metrics very closely. Moreover, as the morphological operators only contain integer operations, our metric holds a very low computational complexity.

6. REFERENCES

[1] Christoph Fehn, "A 3d-tv approach using depth-image-based rendering (DIBR)," in *Proc. of VIIP*, 2003, vol. 3.

[2] Masayuki Tanimoto, Mehrdad Panahpour Tehrani, Toshiaki Fujii, and Tomohiro Yendo, "Free-viewpoint tv," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 67–76, 2011.

[3] Ce Zhu and Shuai Li, "Depth image based view synthesis: New insights and perspectives on hole generation and filling," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 82–93, 2016.

[4] Pierre-Henri Conze, Philippe Robert, and Luce Morin, "Objective view synthesis quality assessment," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 82881M–82881M.

[5] Federica Battisti, Emilie Bosc, Marco Carli, Patrick Le Callet, and Simone Perugia, "Objective image quality assessment of 3d synthesized views," *Signal Processing: Image Communication*, vol. 30, pp. 78–88, 2015.

[6] Dragana Sandić-Stanković, Dragan Kukolj, and Patrick 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.

[7] Dragana Sandić-Stanković, Dragan Kukolj, and Patrick Le Callet, "Multi–scale synthesized view assessment based on morphological pyramids," *Journal of Electrical Engineering*, vol. 67, no. 1, pp. 3–11, 2016.

[8] Emilie Bosc, Philippe Hanhart, Patrick Le Callet, and Touradj Ebrahimi, "A quality assessment protocol for free-viewpoint video sequences synthesized from decompressed depth data," in *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*. IEEE, 2013, pp. 100–105.

[9] Karsten Mueller, Aljoscha Smolic, Kristina Dix, Philipp Merkle, Peter Kauff, and Thomas Wiegand, "View synthesis for advanced 3d video systems," *EURASIP Journal on Image and Video Processing*, vol. 2008, no. 1, pp. 1–11, 2009.

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

[11] Emilie Bosc, Romuald Pepion, Patrick Le Callet, Martin Koppel, Patrick Ndjiki-Nya, Muriel Pressigout, and Luce 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.

[12] Christoph 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.

[13] Alexandru Telea, "An image inpainting technique based on the fast marching method," *Journal of graphics tools*, vol. 9, no. 1, pp. 23–34, 2004.

[14] Yuji Mori, Norishige Fukushima, Tomohiro Yendo, Toshiaki Fujii, and Masayuki Tanimoto, "View generation with 3d warping using depth information for ftv," *Signal Processing: Image Communication*, vol. 24, no. 1, pp. 65–72, 2009.

[15] Patrick Ndjiki-Nya, Martin Köppel, Dimitar Doshkov, Haricharan Lakshman, Philipp Merkle, Karsten Müller, and Thomas Wiegand, "Depth image based rendering with advanced texture synthesis," in *Multimedia and Expo (ICME), 2010 IEEE International Conference on*. IEEE, 2010, pp. 424–429.

[16] Martin Köppel, Patrick Ndjiki-Nya, Dimitar Doshkov, Haricharan Lakshman, Philipp Merkle, Karsten Müller, and Thomas 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.

[17] 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.