

MULTI-SCALE SYNTHESIZED VIEW ASSESSMENT BASED ON MORPHOLOGICAL PYRAMIDS

Dragana Sandić-Stanković*
Dragan Kukolj** — Patrick Le Callet***

The Depth-Image-Based-Rendering (DIBR) algorithms used for 3D video applications introduce geometric distortions affecting the edge coherency in the synthesized images. In order to better deal with specific geometric distortions in the DIBR synthesized images, we propose full-reference metric based on multi-scale pyramid decompositions using morphological filters. The non-linear morphological filters used in multi-scale image decompositions maintain important geometric information such as edges across different resolution levels. We show that PSNR has particularly good agreement with human judgment when it is calculated between detailed images at higher scales of morphological pyramids. Consequently, we propose reduced morphological pyramid peak signal-to-noise ratio metric (MP-PSNR), taking into account only mean squared errors between pyramids' images at higher scales. Proposed computationally efficient metric achieves significantly higher correlation with human judgment compared to the state-of-the-art image quality assessment metrics and compared to the tested metric dedicated to synthesis-related artifacts.

Key words: morphological pyramid, multiscale PSNR, DIBR image quality assessment

1 INTRODUCTION

Depth-Image-Based Rendering (DIBR) techniques can be used to generate views for different 3D video applications: free viewpoint television (FVV), 3DTV, 3D technology based entertainment products and 3D medical applications. The perceptual quality of synthesized views is of great importance for the success of 3D video applications. So far a major effort has been done in view modelling and synthesis. However, lesser effort has been done in developing specific algorithms for assessing the visual quality of synthesized views. The use of subjective tests is expensive, time consuming, cumbersome and practically no feasible in systems where real-time quality score of an image or video sequence is needed. Objective metrics are intended to predict human judgement. The reliability of objective metrics is based on their correlation to subjective assessment results.

Virtual views could be synthesized either from the uncompressed or compressed reference views. The evaluation of DIBR synthesized views from uncompressed data using standard image quality metrics has been discussed in literature [1] for still images from FVV in 2D context. It has been demonstrated that 2D quality metrics originally designed to address image compression distortions are very far to be effective to assess the quality of synthesized views.

The next two metrics have been proposed to improve the performances obtained by standard quality metrics in the evaluation of the DIBR-synthesized images from uncompressed data. Full-reference objective image qual-

ity assessment metric dedicated to view synthesis quality assessment, VSQA, [2] is aimed to handle areas where disparity estimation may fail. It uses 3 visibility maps which characterize complexity in terms of textures, diversity of gradient orientations and presence of high contrast. This metric achieves the gain of 17.8% over SSIM in correlation with subjective measurements. 3DswIM [3] synthesized view image quality metric relies on a comparison of statistical features of wavelet sub-bands of the original and DIBR-synthesized images. Only horizontal detail sub-bands from the first level of Haar wavelet decomposition are used. A registration step and skin detection step are included. A skin detection weights the final quality score in order to penalize distorted blocks containing skin-pixels based on the assumption that a human observer is most sensitive to impairments affecting human subjects. It was reported that 3DswIM outperforms the conventional 2D metrics and tested DIBR image related metrics.

Due to primate visual systems multiscale character [4], multi-resolution image quality assessment methods have improved performances relative to single-resolution methods. Multi-scale structural similarity measure, MS-SSIM [5] is based on linear low-pass pyramid decomposition. Multi-scale image quality measures using information content weighted pooling, IW-SSIM and IW-PSNR [6], use Laplacian pyramid decomposition [7]. CW-SSIM [8] based on multi-orientation steerable pyramid decomposition using multi-scale bandpass oriented filters is simultaneously insensitive to luminance and contrast changes and small geometric distortions of image.

* Institute for Telecommunications and Electronics IRITEL Beograd, Serbia, dragana.sandic@iritel.com ** University of Novi Sad, Faculty of Technical Sciences, Serbia, dragan.kukolj@rt-rk.uns.ac.rs *** Ecole polytechnique de l'Universite de Nantes, IRCCyN Lab, France, patrick.lecallet@univ-nantes.fr

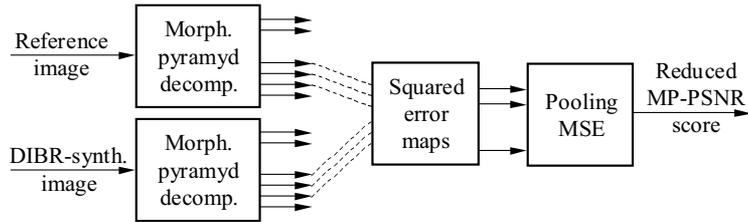


Fig. 1. Morphological multiscale image quality assessment framework

DIBR algorithms introduce new types of artefacts mostly located around disoccluded regions. The introduced geometric distortions are affecting edges coherency in the synthesized images. Typical rendering errors include black holes, boundary blur, edge displacements or misalignments. These artefacts are consequently challenging for standard quality metrics, usually tuned for other distortion types.

In order to better deal with specific geometric distortions in DIBR synthesized images, we propose multi-scale image quality assessment metric based on morphological filters in multiresolution image decomposition. Introduced non-linear morphological filters maintain important geometric information such as edges across different resolution levels [9]. In the previous work [10], we explored morphological wavelet decompositions for the multiscale metric MW-PSNR, which achieves much higher correlation with human judgment compared to the state-of-the-art image quality measures. We also explored morphological pyramid decomposition in the multiscale metric MP-PSNR [11] which uses mean squared errors of all pyramids' sub-bands. There is inherent congruence between the morphological pyramid decomposition scheme and human visual perception [12]. In this paper, we show that the performances of PSNR calculated between detail images at higher scales of two morphological band-pass pyramids, the reference image pyramid and the DIBR-synthesized image pyramid, show much better agreement with human judgment than the performances of PSNR calculated on lower pyramids' scales. Therefore, we propose the improved version of morphological pyramid-based measure, reduced morphological pyramid-peak signal-to-noise ratio, MP-PSNR, using only detail images from higher pyramids' scales for the calculation of MSE. After the morphological bandpass pyramid decomposition of the reference and the synthesized images, we propose to calculate squared error maps between the appropriate detail images only at higher pyramids' scales of the two pyramids. Mean squared errors from higher pyramids' scales are combined into the multi-scale mean squared error, which is transformed into reduced multi-scale peak signal-to-noise ratio measure. Reduced MP-PSNR is highly correlated with the judgment of human observers, much better than standard IQA metrics. It shows much better performances than tested metric dedicated to synthesis-related artefacts. The performances of reduced morphological multi-scale measure are improved

comparing to its full-scales version. Since the morphological operators involve only integers and only max, min and addition in their computation, as well as simple calculation of MSE, the proposed morphological multi-scale metric is of low computational complexity. Reduced MP-PSNR has lower computational complexity than its full version. The distortions introduced by view synthesis algorithms are evaluated using IRCCyN/IVC DIBR images dataset [13].

In the next section, the proposed multi-scale metric, the reduced MP-PSNR is described. Description of morphological pyramid decomposition, the distortion computation stage and pooling stage of the reduced MP-PSNR are given in Section 2. The performances of the proposed metric and discussion are presented in Section 3, while the conclusion is given in Section 4.

2 PROPOSED MORPHOLOGICAL MULTI-SCALE METRIC

Mathematical morphology is a set-theoretic method for image analysis which provide a quantitative description of geometric structure of an image. It considers images as sets which permits geometry-oriented transformations of the images. For the multiscale image decomposition we have used basic morphological operators dilation (D) and erosion (E), [14, 15]:

$$D: \quad dilation_{SE}(f)(x) = \max_{y \in SE} f(x - y), \quad (1)$$

$$E: \quad erosion_{SE}(f)(x) = \min_{y \in SE} f(x + y), \quad (2)$$

where f is a grayscale image and SE is binary (flat) structuring element. Dilation/erosion are calculated as the maximum/minimum in the neighbourhood defined by structuring element.

A multi-scale representation is completely specified by the transformation from a finer scale to a coarser scale. In linear scale-spaces the operator for changing scale is a convolution by a Gaussian kernel. After the convolution with Gaussian kernel the images are uniformly blurred, also the regions of particular interest like the edges [16]. To overcome this issue, nonlinear multiresolution signal decomposition schemes based on morphological operators have been proposed to maintain edges through scales [9].

The proposed multi-scale metric for DIBR image quality assessment, reduced morphological pyramid peak

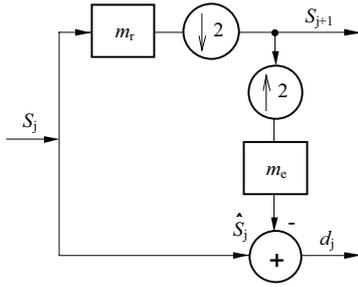


Fig. 2. Morphological bandpass pyramid decomposition scheme: m_r morphological reduce operator followed by down-sampling, m_e morphological expand operator preceded by up-sampling

signal-to-noise ratio measure (MP-PSNR), can be described as three-stage process, Fig. 1. In the first stage both the reference and the distorted images are decomposed into a set of lower resolution images using multi-resolution morphological pyramid decomposition. We propose to use nonlinear morphological operators in the multi-scale pyramid decomposition in order to better deal with specific geometric distortions in DIBR synthesized images. Introduced non-linear morphological filters maintain important geometric information such as edges across different resolution levels [9]. In the second stage of the morphological multi-scale IQA framework, Fig. 1, squared error maps are evaluated using only the detail images at higher scales of the two morphological pyramids, the reference image pyramid and the DIBR-synthesized image pyramid. In the third stage, a pooling is employed to convert these maps into a single quality scores, mean squared errors, and finally these scores are combined into the multi-scale image quality measure, multi-scale mean squared error, which is transformed into the reduced morphological pyramid peak signal-to-noise ratio measure, MP-PSNR.

We propose to use morphological bandpass pyramid (MBP) decomposition in the first stage of morphological multi-scale IQA framework. Morphological bandpass pyramid is generated using the Laplacian type pyramid decomposition scheme [7]. Single level of morphological bandpass pyramid decomposition is shown on Fig. 2, where m_r is morphological reduce filter for low-pass filtering in analysis step and m_e is morphological expand filter for interpolation filtering in synthesis step. Morphological bandpass pyramid with M decomposition levels consists of detail (error) images of decreasing size d_j , $j = 0, \dots, M-1$ and the coarse lowest resolution image s_M [12]. MBP tends to enhance image features such

as edges which are segregated by scale in the various pyramid levels. Enhanced features are segregated by size: fine details are prominent in the lower level images while progressively coarser features are prominent in the higher level images.

The proposed morphological bandpass pyramid is morphological pyramid adjunction, MBP ED [16], which uses morphological operator erosion as reduce operator m_r followed by down-sampling, Fig. 2, and morphological operator dilation as expand operator m_e preceded by up-sampling. MBP ED pyramid based on adjunction satisfies the property that the detail signal is always nonnegative. At any scale change, maximum luminance at the coarser scale is always lower than the maximum luminance at the finer scale, the minimum is always higher. MBP ED pyramid using structuring element of size 2×2 is morphological Haar pyramid [17]. Symmetrized version of the morphological Haar pyramid MBP ED uses structuring element of size 3×3 . A lower resolution image s_{j+1} of the MBP ED pyramid representation, Fig. 2, is obtained by applying morphological operator erosion on the previous pyramid level image s_j and down-sampling it by factor 2 on each image dimension (δ^\downarrow). The eroded image at level j , s_{jE} with the square structuring element of size $(2r+1) \times (2r+1)$ is calculated as

$$s_{jE}(m, n) = \min\{s_j(m+k, n+l) \mid -r \leq k, l \leq r\},$$

$$s_{j+1} = \delta^\downarrow(s_{jE}). \quad (3)$$

A detail image is derived by subtracting from each level an interpolated version of the next coarser level. The image s_{j+1} of the next pyramid level is up-sampled by factor 2 on each dimension (δ^\uparrow) leading to the image s_{jU} . Morphological operator dilation is applied on the up-sampled image s_{jU} to produce expanded image \hat{s}_j . The detail image d_j is obtained as the difference of the low-pass pyramid image s_j and expanded image from the next level \hat{s}_j

$$s_{jU} = \delta^\uparrow(s_{j+1}),$$

$$\hat{s}_j(m, n) = \max\{s_{jU}(m-k, n-l) \mid -r \leq k, l \leq r\}, \quad (4)$$

$$d_j = s_j - \hat{s}_j.$$

Using square structuring element, morphological reduce and expand filtering is implemented more efficiently separately by rows and columns using the structuring elements of size $1 \times (2r+1)$ for rows and $(2r+1) \times 1$ for columns.

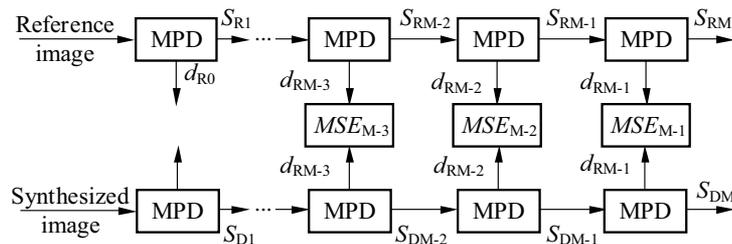


Fig. 3. Reduced MP-PSNR is based on MSE between pyramids' detail images at higher scales



Fig. 4. Frames from multi-view video sequences: Book Arrival, Lovebird1 and Newspaper

In the second stage of the proposed multi-scale IQA framework, squared error maps are calculated between the detail images from higher scales of the two pyramids: the reference image pyramid and the DIBR-synthesized image pyramid. From the squared error maps, mean squared errors are calculated and combined into the multi-scale mean squared error which is transformed into multi-scale peak signal-to-noise ratio in the third stage of the multi-scale IQA framework.

Mean squared error (MSE) and peak signal-to-noise ratio (PSNR) are simple and widely used full-reference objective image distortion/quality metrics. The MSE remains the standard criterion for the assessment of signal quality and fidelity. It has a clear physical meaning — it is the natural way to define the energy of the error signal. The MSE is an excellent metric in the context of optimization. Moreover, through the literature, competing algorithms have most often been compared using MSE/PSNR [18]. In this paper, MSE is proposed for distortion measurement of the pyramids' bandpass images. At pyramids' scale j , MSE_j is calculated as the mean value of the squared error map between the reference pyramid image d_{Rj} and the distorted pyramid image d_{Dj} of size $K_j N_j$

$$MSE_j = \frac{1}{N_j K_j} \sum_{k=1}^{K_j} \sum_{n=1}^{N_j} (d_{Rj}(k, n) - d_{Dj}(k, n))^2. \quad (5)$$

We propose to calculate multi-scale reduced pyramid mean squared error MP_MSE as the average of MSE_j values at three highest pyramids' scales, Fig. 3

$$MP_MSE = \sum_{j=M-3}^{M-1} \beta_j MSE_j, \quad (6)$$

where equal value weights $\beta_j = \frac{1}{3}$ are used and M is the number of decomposition levels. Finally, MP_MSE is transformed into reduced morphological pyramid peak signal-to-noise MP_PSNR , where R is the maximum dynamic range of the image.

$$MP_PSNR = 10 \log_{10} \frac{R^2}{MP_MSE}. \quad (7)$$

Matlab implementation of MP_PSNR is available online [20].

3 RESULTS AND DISCUSSION

The performances of the reduced MP_PSNR are evaluated using IRCC_yN/IVC DIBR images database [13] introduced in [1]. It contains frames from 3 multi-view video sequences: Book Arrival (1024×768 , 16 cameras with 6.5 cm spacing), Lovebird1 (1024×768 , 12 cameras with 3.5 cm spacing) and Newspaper (1024×768 , 9 cameras with 5 cm spacing). The frames from these sequences are shown on Fig. 4. The selected contents are representative and used by MPEG also.

For each sequence four virtual views are generated on the positions corresponding to those positions obtained by the real cameras using seven depth image based rendering algorithms, named A1–A7 [21–25]. One key frame from each synthesized sequence is randomly chosen for the DIBR images database [1]. For these key frames subjective assessment in form of mean opinion scores (MOS) is provided. The difference MOS (DMOS) is calculated as the difference between the reference frame's MOS and the synthesized frame's MOS. In the algorithm A1 [21], the depth image is pre-processed by a low-pass filter, borders are cropped and the image is interpolated to reach its original size. Due to very noticeable object shifting artefacts in the frames generated by algorithm A1, these frames are excluded from the experiment. The focus remains on images synthesized using A2–A7 DIBR algorithms and without registration procedure for image alignment. The algorithm A2 is based on A1 except that the borders are not cropped but in-painted. The algorithm A3 [22] use in-painting method to fill in the missing parts in the virtual image which introduces blur in the disoccluded area. This algorithm was adopted as the reference software for MPEG standardization experiments in 3D Video group. The algorithm A4 performs hole-filling method aided by depth information [23]. The algorithm A5 uses a patch-based texture synthesis as the hole-filling method [24]. The algorithm A6 uses depth temporal information to improve synthesis in the disoccluded areas [25]. The frames generated by algorithm A7 contain unfilled holes.

To compare the performances of the image quality measures the following evaluation metrics are used: root mean squared error between the subjective and objective scores (RMSE), Pearsons correlation coefficient with

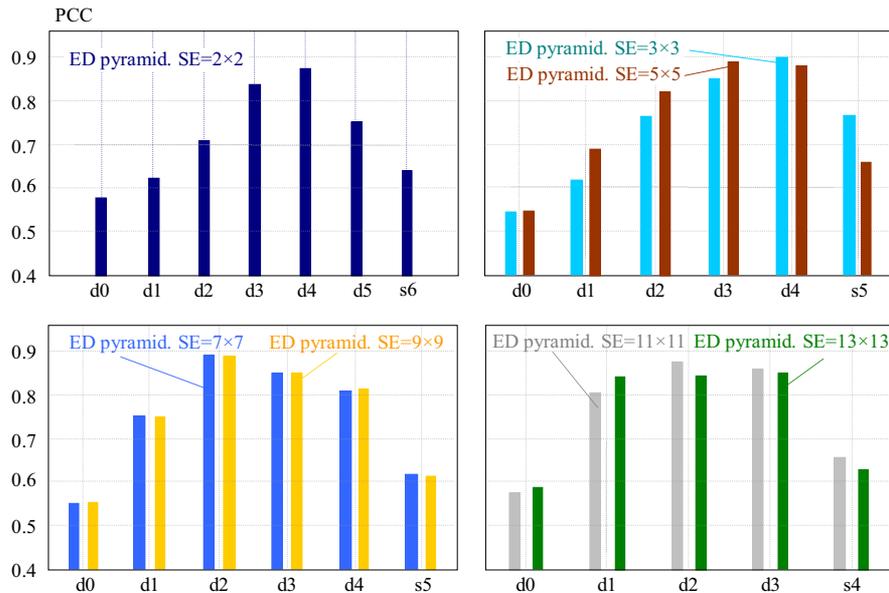


Fig. 5. Pearson's correlation coefficients of pyramid images' PSNR versus DMOS

Table 1. The best PSNR performances by pyramid scales for SE of different sizes. The performances of PSNR calculated between the reference images and the synthesized images are presented in the last row

SE	scale	image	RMSE	PCC	SCC
2×2	5	d4	0.3270	0.8792	0.8147
3×3	5	d4	<u>0.3076</u>	<u>0.8939</u>	<u>0.8671</u>
5×5	4	d3	0.3130	0.8899	0.8656
7×7	3	d2	0.3180	0.8862	0.8485
9×9	3	d2	0.3239	0.8816	0.8697
11×11	3	d2	0.3307	0.8763	0.8513
13×13	4	d3	0.3597	0.8517	0.7859
...	...	s0	0.4525	0.7519	0.6766

nonlinear mapping between the subjective scores and objective measures (PCC) and Spearman's rank order correlation coefficient (SCC). The calculation of DMOS from given MOS and nonlinear mapping between the subjective scores and objective measures are done according to the HDTV video quality evaluation models defined by VQEG HDTV group [26].

We have investigated PSNR performances by pyramid scales, between the appropriate pyramids' images at all pyramid scales. The reference image and the DIBR synthesized image are decomposed into a set of lower resolution pyramid images using morphological band-pass pyramid decomposition with erosion/dilation as reduce/expand operation. At each pyramid scale, PSNR is calculated between the appropriate detail images of the two pyramids.

In order to find out how the PSNR is correlated to subjective opinion by pyramid scales, PCC coefficients of PSNR versus DMOS are calculated at all pyramid scales. The tests are performed using SE s of different sizes for morphological filtering in MBP ED pyramid decomposition. Pearson's correlation coefficients of PSNR versus DMOS for pyramid images at all pyramid scales using SE s of different sizes are shown on Fig. 5. From these bar charts, it can be concluded that the objective quality of

an image is heavily dependent upon the scale at which the image is analyzed. The Pearson's correlation coefficients of PSNR versus DMOS for lower resolution detail images on higher pyramid scales are much better than for the higher resolution images on lower pyramid scales. Also, PSNR performances on higher pyramid scales are much better than the PSNR performances for the case when the PSNR is calculated between the original and the DIBR synthesized images in the base band. For the morphological pyramid decomposition using SE of size 2×2 and 3×3 , the best PSNR performances are obtained for the case when the PSNR is calculated at pyramid scale 5 between detail images d_4 (Fig. 5, two upper charts). Using the morphological pyramid decomposition with SE of size 5×5 the best PSNR performances are obtained at pyramid scale 4 between detail images d_3 (Fig. 5, the upper right chart), while for the pyramid decomposition with larger SE , the best PSNR performances are obtained at pyramid scale 3 for detail images d_2 , (Fig. 5, two lower charts).

The best PSNR performances by pyramid scales for different sizes of SE used in morphological pyramid decomposition are shown in Table 1. For the morphological pyramid decomposition using SE of size 3×3 , the best PSNR performances are achieved at pyramid scale 5, Pearson correlation coefficient 0.89 and Spearman correlation coefficient 0.867. It is shown in [18, 19] that MSE shows poor performances in some cases when it is used as single scale metric on the full resolution base band images. In this paper, it is demonstrated that PSNR shows very good agreement with human quality judgments when it is calculated for lower resolution detail images on higher scales of morphological bandpass pyramids.

Based on the results of PSNR performances calculated separately by pyramid scales, it can be concluded that the PSNR performances of lower resolution detail images on higher pyramid scales are much better than the PSNR performances of higher resolution detail images on lower

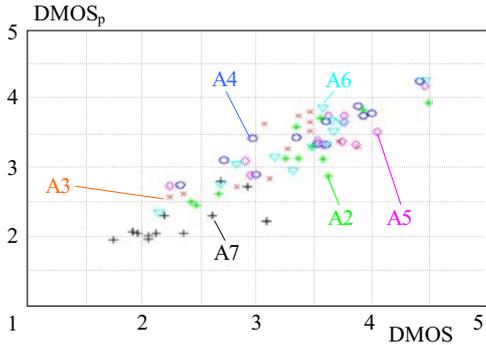


Fig. 6. Fitted scores of reduced MP-PSNR versus DMOS

Table 2. Performances of the reduced and full versions of MP-PSNR

reduced MP-PSNR				
SE	levels	RMSE	PCC	SCC
2×2	4–6	0.3660	0.8459	0.7775
3×3	3–5	0.3252	0.8806	0.8185
5×5	3–5	<u>0.2936</u>	<u>0.9039</u>	<u>0.8634</u>
7×7	3–5	0.2931	0.9042	0.8573
9×9	2–4	0.2997	0.8996	0.8614
11×11	2–4	0.2922	0.9048	0.8684
13×13	2–4	0.2920	0.9050	0.8684

full MP-PSNR				
SE	levels	RMSE	PCC	SCC
2×2	6	0.4101	0.8019	0.7083
3×3	5	0.3996	0.8131	0.7101
5×5	5	0.3561	0.8549	0.7759
7×7	5	0.3264	0.8796	0.8050
9×9	4	0.3319	0.8753	0.8014
11×11	4	0.3165	0.8874	0.8175
13×13	4	0.3221	0.8830	0.8021

pyramid scales. Therefore, we propose reduced version of MP-PSNR using only lower resolution detail images from the higher pyramid scales. Reduced MP-PSNR is calculated as the average MSE of the three highest scales detail images (6). The impact of structuring element's size used in morphological operations in MBP ED pyramid decompositions on MP-PSNR performances is investigated. The performances of reduced MP-PSNR are calculated using MBP ED pyramid decomposition with square SEs of different sizes from 2×2 to 13×13 . More features are removed from the image as larger SE is used. The performances of the reduced MP-PSNR are presented in the Table 2, top. For the MBP ED pyramid decomposition using SE of size 2×2 , the reduced version of MP-PSNR is calculated using only detail images from scales 4–6. For the MBP ED pyramid decompositions using SEs of size 3×3 , 5×5 and 7×7 the reduced version of MP-PSNR is calculated using only detail images from scales 3–5. For the MBP ED pyramid decomposition using SE of size 9×9 , 11×11 and 13×13 , the reduced version of MP-PSNR is calculated using only detail images from levels 2–4. The best result is obtained with reduced MP-PSNR using MBP ED pyramid with SE of size 5×5 (Pearson's

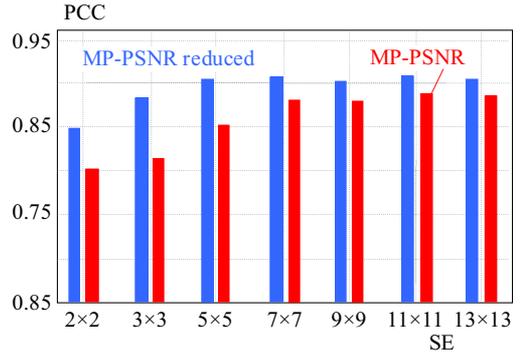


Fig. 7. Pearson's correlation coefficients of the reduced and full MP-PSNR

90.39%, Spearman 86.3%). Scatter plot of nonlinearly mapped reduced MP-PSNR versus DMOS for that case is shown in Fig. 6. Each point represents one frame from the DIBR database.

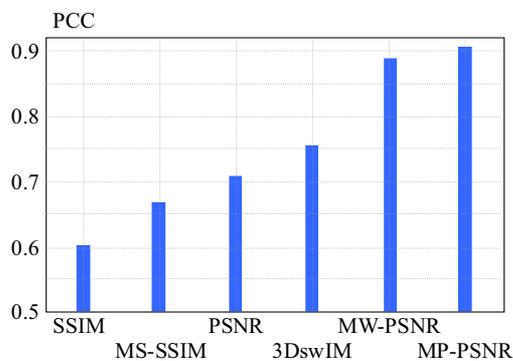
The performances of the full version of multi-scale metric MP-PSNR using all pyramids' images presented in [11] are shown in the Table 2, bottom. The number of decomposition levels for the best MP-PSNR performances depends on the size of SE [11]. The highest correlation of the full version of MP-PSNR with subjective scores, Pearson 0.887 and Spearman 0.817, is achieved using MBP ED pyramid decomposition with SE of size 11×11 pixels. The comparison of Pearson's correlation coefficients of the full and the reduced MP-PSNR versus DMOS for different sizes of SE is shown on Fig. 7. Reduced MP-PSNR has better performances than its full version for SE of all sizes. For instance, Pearson's correlation coefficient of reduced MP-PSNR using SE of size 3×3 is 6.75% higher than Pearson's correlation coefficient of full MP-PSNR. Reduced MP-PSNR is computationally more efficient than its full version as the mean squared errors are calculated only for lower resolution pyramid images.

3.1 The results summary

The performances of the the selected proposed metrics, the commonly used 2D image quality assessment metrics and the metric dedicated to synthesis-related artefacts, 3DswIM [3], are presented in Table 3. The considered commonly used 2D metrics are: PSNR, universal quality index UQI [27], structural similarity index SSIM [28], multi-scale structural similarity MS-SSIM [5], information weighted IW-PSNR [6] and IW-SSIM [6]. Single-scale structural similarity SSIM is calculated between the original and the synthesized images using the matlab code [29]. 3DswIM is calculated using the matlab code [30]. Selected versions of the proposed metrics using morphological pyramid decompositions presented in Table 3 are: PSNR calculated on scale 5 of the MBP ED pyramid using SE of size 3×3 ; reduced and full versions of MP-PSNR using SE of size 5×5 in pyramid MBP ED decomposition. Selected versions of the metrics based on morphological wavelet decompositions [10] are: PSNR calculated on scale 6 between vertical

Table 3. Performances of the selected proposed metrics and other metrics

	RMSE	PCC	SCC
proposed metrics			
PSNR(d4), $SE = 3 \times 3$	0.3076	0.8939	0.8671
<u>reduced MP-PSNR</u> $SE = 5 \times 5$	<u>0.2936</u>	<u>0.9039</u>	<u>0.8634</u>
full MP-PSNR $SE = 5 \times 5$	0.3561	0.8549	0.7759
PSNR(d61) using minLift wavelet [10]	0.3167	0.8872	0.8281
reduced MW-PSNR [10]	0.3188	0.8855	0.8298
full MW-PSNR [10]	0.3565	0.8545	0.7750
metric dedicated to DIBR synthesized images			
3DswIM [3][30]	0.4868	0.7049	0.6396
commonly used 2D IQA metrics			
PSNR	0.4525	0.7519	0.6766
IW-PSNR [6]	0.5267	0.6411	0.5320
UQI [27]	0.5199	0.6529	0.5708
SSIM [28][29]	0.5513	0.5956	0.4424
MS-SSIM [5]	0.5127	0.6649	0.5188
IW-SSIM [6]	0.5350	0.6265	0.4856

**Fig. 8.** Pearson's correlation coefficients of the proposed metric and other metrics versus DMOS

wavelet sub-bands of the two wavelet image representations using morphological min Lift wavelet; reduced and full versions of MW-PSNR using morphological wavelet decomposition with min Haar wavelet. The performances of the proposed metrics are much better than the performances of the commonly used 2D metrics and much better than the performances of the synthesis-related metric, 3DswIM. The Pearson's correlation coefficients of the selected commonly used 2D metrics, 3DswIM, and the reduced versions of MP-PSNR and of MW-PSNR are shown on Fig. 8.

3.2 Computational complexity of the reduced MP-PSNR

Since the morphological operators used in morphological multiresolution decomposition schemes involve only integers and only max, min and addition in their computation the calculation of morphological pyramid has low computational complexity. The calculation of MSE is of low computational complexity, 1 subtraction, 1 multiplication and 1 addition are needed per pixel. In case of reduced MP-PSNR, MSE is calculated only for lower resolution detail images at higher pyramid scales. Therefore, the calculation of reduced MP-PSNR is not computationally demanding.

Morphological reduce and expand filtering in the morphological pyramid decomposition scheme is implemented with lower computational complexity in two steps, separably by columns and rows using SEs of size $P \times 1$ pixels for columns and $1 \times P$ pixels for rows, $P = 2, 3, 5, 7, 9, 11, 13$. In the reduce filtering, down-sampling is performed after filtering by columns and after filtering by rows. In the expand filtering, up-sampling is performed before each filtering step, before filtering by rows and before filtering by columns. For morphological reduce filtering of the image with NK pixels, using operator erosion with SE of size $P \times 1$ by columns and $1 \times P$ by rows, $\frac{3}{2}(P-1)NK$ comparisons (C) are needed. The same number of comparisons are needed for interpolation expand filtering by morphological operator dilation. The pyramid's detail image is calculated as the difference of the low-pass pyramid image and the expanded image of the next pyramid's scale using operator subtraction (S). The number of operations for the calculation of the detail image at first pyramid scale is $(3(P-1) \cdot C + S) \cdot NK$. The total number of operations for the calculation of morphological bandpass erosion/dilation pyramid in L levels is

$$\begin{aligned} & (3(P-1)C + S)NK \left(1 + \frac{1}{2^2} + \frac{1}{2^4} + \dots + \frac{1}{2^{2(L-1)}} \right) \\ & = (3(P-1)C + S) \frac{4}{3} NK, \quad L \rightarrow \infty. \quad (8) \end{aligned}$$

The number of operations for the calculation of morphological bandpass ED pyramid depends on the size of $SE(P)$ as it is shown in Table 4. With the increase of structuring element's size, the computational complexity increases.

The number of operations for the calculation of MSE for the reduced version of MP-PSNR is shown in the upper part of Table 5 and for the full version of MP-PSNR in the bottom part of Table 5. The computational complexity of the reduced MP-PSNR is smaller than the computational complexity of the full MP-PSNR because mean

Table 4. The number of operations for the calculation of the morphological bandpass erosion/dilation pyramid. The image of size NK is decomposed in L levels using SE of size $1 \times P$ pixels by rows and $P \times 1$ pixels by columns. Operations subtraction (S), comparison (C) are used

$P(SE)$	L	operations number for decomp.
2	6	$(3C + S) \frac{1365}{1024} NK$
3	5	$(6C + S) \frac{341}{256} NK$
5	5	$(12C + S) \frac{341}{256} NK$
7	5	$(18C + S) \frac{341}{256} NK$
9	4	$(24C + S) \frac{85}{64} NK$
11	4	$(30C + S) \frac{85}{64} NK$
13	4	$(36C + S) \frac{85}{64} NK$

Table 5. The number of operations for the calculation of MSE for the reduced and full MP-PSNR. MBP ED pyramid decomposition is performed in L levels using SE of size $P \times P$. MSE is calculated using subtraction (S), multiplication (M) and addition (A)

$P(SE)$	L	operations number for MSE
reduced MP-PSNR		
2	4-6	$(S + M + A) \frac{21}{1024} NK$
3	3-5	$(S + M + A) \frac{21}{256} NK$
5	3-5	
7	3-5	
9	2-4	$(S + M + A) \frac{21}{64} NK$
11	2-4	
13	2-4	
full MP-PSNR		
2	6	$(S + M + A) \frac{1365}{1024} NK$
3	5	$(S + M + A) \frac{341}{256} NK$
5	5	
7	5	
9	4	$(S + M + A) \frac{84}{64} NK$
11	4	
13	4	

squared errors are calculated only for lower resolution detail images at three highest pyramid scales for reduced MP-PSNR while for the full MP-PSNR, mean squared errors are calculated at all pyramid scales.

4 CONCLUSION

Most of the Depth Image Based Rendering (DIBR) techniques produce images which contain non-uniform geometric distortions affecting the edge coherency. This type of distortions are challenging for common image quality assessment metrics. We propose full-reference metric based on morphological multi-scale pyramid decomposition in order to better deal with specific geometric distortions in the DIBR synthesized images. Introduced non-linear morphological filters maintain important geometric information such as edges across different resolution scales. The proposed metric measures the edge distortion using MSE between the detail images from the

higher scales of two morphological pyramids: the reference image pyramid and the DIBR synthesized image pyramid.

We show that PSNR has particularly good agreement with human judgment when it is calculated between the appropriate detail images at higher decomposition scales of the two morphological pyramid image representations. The tests are performed using IRCCyN/IVC DIBR images database. PSNR calculated on scale 5 of the MBP ED pyramid using square structuring element of size 3×3 has very good performances (Pearson's 0.894, Spearman 0.867). Reduced MP-PSNR achieves high improvement of performances over standard image quality metrics and over tested metric dedicated to synthesis-related artefacts. Reduced version of MP-PSNR using morphological pyramid decomposition MBP ED with square structuring element of size 5×5 achieves the improvement 15.2% of correlation over PSNR (Pearson's 0.904, Spearman 0.863). It has also better performances and lower computational complexity than the full version of MP-PSNR. Reduced MP-PSNR using MBP ED with square structuring element of size 5×5 achieves the improvement 4.9% of correlation over the full version of MP-PSNR. Since the morphological operators involve only integers and only min, max and addition in their computation, as well as simple calculation of MSE, reduced MP-PSNR is computationally efficient procedure.

Acknowledgement

This work was partially supported by COST Action IC1105-3D ConTourNet, the Ministry of Education, Science and Technological Development of the Republic of Serbia under Grant TR-32034 and by the Secretary of Science and Technology Development of the Province of Vojvodina under Grant 114-451-813/2015-03.

REFERENCES

- [1] BOSC, E.—PEPION, R.—Le CALLET, P.—KOPPEL, M.—NDJIKI-NYA, P.—PRESSIGOUT, M.—MORIN, L.: Towards a New Quality Metric for 3-D Synthesized View Assessment, IEEE Journal on Selected Topics in Signal Processing **05** No. 07 (2011), 1332–1343.
- [2] CONZE, P. H.—ROBERT, P.—MORIN, L.: Objective View Synthesis Quality Assessment, Proc. SPIE 8288, Stereoscopic Displays and Applications XXIII (2012).
- [3] BATTISTI, F.—BOSC, E.—CARLI, M.—Le CALLET, P.—PERUGIA, S.: Objective Image Quality Assessment of 3D Synthesized Views, Elsevier Signal Processing: Image Communication **30** No. 01 (2015), 78–88.
- [4] ADELSON, E.—SIMONCELLI, E.—FREEMAN, W.: Pyramids and Multiscale Representations, Proc. European Conference on Visual Perception, Paris, Aug 1990.
- [5] WANG, Z.—SIMONCELLI, E.—BOVIK, A. C.: Multi-Scale Structural Similarity for Image Quality Assessment, Asilomar Conference on Signals, Systems and Computers, Nov 2003.
- [6] WANG, Z.—LI, Q.: Information Content Weighting for Perceptual Image Quality Assessment, IEEE Trans. on Image Processing **20** No. 05 (2011), 1185–1198.

- [7] BURT, P. J.—ADELSON, E. H.: The Laplacian Pyramid as a Compact Image Code, *IEEE Trans. on Communications* **31** No. 04 (1983), 532–540.
- [8] WANG, Z.—SIMONCELLI, E.: Translation Insensitive Image Similarity in Complex Wavelet Domain, *Proc. IEEE International Conference on Acoustics, Speech and Signal processing*, 2005, pp. 573–576.
- [9] MARAGOS, P.—SCHAFER, R.: Morphological Systems for Multidimensional Signal Processing, *Proceedings of the IEEE* **78** No. 04 (1990), 690–710.
- [10] SANDIC-STANKOVIC, D.—KUKOLJ, D.—Le CALLET, P.: DIBR Synthesized Image Quality Assessment based on Morphological Wavelets, *International Workshop on Quality of Multimedia Experience QoMEX, Costa Navarino, Greece*, 2015.
- [11] SANDIC-STANKOVIC, D.—KUKOLJ, D.—Le CALLET, P.: DIBR Synthesized Image Quality Assessment based on Morphological Pyramids, *3DTV-CON Immersive and interactive 3D Media Experience over Networks, Lisbon*, 2015.
- [12] TOET, A.: A morphological pyramidal image decomposition, *Pattern Recognition Letters* **09** No. 04 (1989), 255–261.
- [13] ftp://ftp.ivc.polytech.univ-nantes.fr/IRCCyN_IVC_DIBR_Images
IRCCyN/IVC DIBR images database.
- [14] STERNBERG, S.: Gray Scale Morphology, *Journal on Computer Vision, Graphics and Image Processing* **35** No. 03 (1986), 335–355.
- [15] HARALICK, R.—STERNBERG, S.—ZHUANG, X.: Image Analysis using Mathematical Morphology, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **09** No. 04 (1987), 532–550.
- [16] MEYER, F.—MARAGOS, P.: Nonlinear Scale-Space Representation with Morphological Levelings, *Journal of Visual Communication and Image Representation* **11** No. 02 (2000), 245–265.
- [17] GOUTSIAS, J.—HEIJMANS, H.: Nonlinear Multiresolution Signal Decomposition Schemes — Part I: Morphological Pyramids, *IEEE Trans. on Image Processing* **09** No. 11 (2000), 1862–1876.
- [18] WANG, Z.—BOVIK, A.: Mean Squared Error: Love it or Leave it, *IEEE Signal Processing Magazine* **26** No. 01 (2009), 98–117.
- [19] WANG, Z.—BOVIK, A.—LU, L.: Why is Image Quality Assessment so Difficult, *IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Orlando, FL, US*, 2002, pp. 3313–3316.
- [20] <https://sites.google.com/site/draganasandicstankovic/code/mp-psnr> Matlab code for MP-PSNR.
- [21] FEHN, C.: Depth Image based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV, *Proc. SPIE, Stereoscopic Displays and Applications XV, San Jose, CA*, 2004, pp. 93–104.
- [22] MORI, Y.—FUKUSHIMA, N.—YENDO, T.—FUJII, T.—TANIMOTO, M.: View Generation with 3D Warping using Depth Information for FTV, *Signal Processing: Image Communication* **24** No. 1-2 (2009), 65–72.
- [23] MULLER, K.—SMOLIC, A.—DIX, K.—MERKLE, P.—KAUFF, P.—WIEGAND, T.: View Synthesis for Advanced 3D Video Systems, *EURASIP Journal on image and video processing* (2008), doi:10.1155/2008/438148.
- [24] NDJIKI-NYA, P.—KOPPEL, P.—DOSHKOV, M.—LAKSHMAN, H.—MERKLE, P.—MULLER, K.—WIEGANDT.: Depth Image based Rendering with Advanced Texture Synthesis, *IEEE Int. Conf. on Multimedia & Expo* (2010), 424–429, Suntec City.
- [25] KOPPEL, M.—NDJIKI-NYA, P.—DOSHKOV, M.—LAKSHMAN, H.—MERKLE, P.—MULLER, K.—WIEGAND, T.: Temporally Consistent Handling of Disocclusions with Texture Synthesis for Depth-Image-Based Rendering, *IEEE Int. Conf. on Image Processing, Hong Kong*, 2010, pp. 1809–1812.
- [26] VQEG HDTV Group, Test Plan for Evaluation of Video Quality Models for Use with High Definition TV Content. 2009.
- [27] WANG, Z.—BOVIK, A. C.: A Universal Image Quality Index, *IEEE Signal Processing Letters* **09** No. 03 (2002), 81–84.
- [28] WANG, Z.—BOVIK, A. C.—SHEIKH, H. R.—SIMONCELLI, E.: Image Quality Assessment: From Error Visibility to Structural Similarity, *IEEE Trans. on Image Processing* **13** No. 04 (2004), 600–612.
- [29] https://ece.uwaterloo.ca/~z70wang/research/ssim/ssim_index.m Matlab code for SSIM.
- [30] <http://www.comlab.uniroma3.it/3DSwIM.html> Matlab code for 3DSwIM.

Received 7 September 2015

Dragana Sandic-Stankovic received her Diploma degree in electronics in 1991, from the High military-technical school of land force of SFR Yugoslavia, MSc degree in digital data transmission in 1999, from the Faculty of Electrical Engineering, University of Belgrade, Serbia. She is a Software Developer at the Institute for Telecommunications and Electronics, IRITEL, Belgrade. Her research interests include multiresolution decompositions, morphological filters, image processing. She is an author of about 40 conference papers.

Dragan Kukolj received his Diploma degree in control engineering in 1982, MSc degree in computer engineering in 1988, and PhD degree in control engineering in 1993, all from the University of Novi Sad, Serbia. He is currently a Professor of computer-based systems with Department of Computing and Control, Faculty of Engineering, University of Novi Sad. His main research interests include soft computing, data mining techniques and computer-based systems integration with applications in video processing and digital signal processing. He has published about 190 papers in referred journals and conference proceedings, 5 books and he is co-inventor of 12 national patents and patent applications and the founder and coordinator of Intellectual Property Center of University of Novi Sad.

Patrick Le Callet received both an MSc and a PhD degree in image processing from Ecole polytechnique de l'Université de Nantes. Since 2003 he teaches at Ecole polytechnique de l'Université de Nantes in the Electrical Engineering and the Computer Science departments where is now a Full Professor. Since 2006, he is the head of the Image and Video Communication lab at CNRS IRCCyN. He is mostly engaged in research dealing with the application of human vision modeling in image and video processing. His current centers of interest are 3D image and video quality assessment, watermarking techniques and visual attention modeling and applications. He is co-author of more than 200 publications and communications and co-inventor of 13 international patents on these topics. He also co-chairs within the VQEG the “HDR Group” and “3DTV” activities. He is currently serving as associate editor for *IEEE Trans. on Image Processing*, *IEEE Trans. on Circuit System and Video Technology*, *EURASIP Journal on Image and Video Processing*, and *SPIE Electronic Imaging*.