Quality Assessment of DIBR-Synthesized Views Based on Sparsity of Difference of Closings and Difference of Gaussians

Dragana Sandic-Stankovic, Dragan Kukolj, Patrick Le Callet

Quality Assessment of DIBR-Synthesized Views Based on Sparsity of Difference of Closings and Difference of Gaussians

Dragana D. Sandic´-Stankovi´c, Dragan D. Kukolj, *Senior Member, IEEE*, and Patrick Le Callet, *Fellow, IEEE*

***Abstract*—Images synthesized using depth-image-basedrendering (DIBR) techniques may suffer from complex structural distortions. The goal of the primary visual cortex and other parts of brain is to reduce redundancies of input visual signal in order to discover the intrinsic image structure, and thus create sparse image representation. Human visual system (HVS) treats images on several scales and several levels of resolution when perceiving the visual scene. With an attempt to emulate the properties of HVS, we have designed the no-reference model for the quality assessment of DIBR-synthesized views. To extract a higher-order structure of high curvature which corresponds to distortion of shapes to which the HVS is highly sensitive, we define a morphological oriented Difference of Closings (DoC) operator and use it at multiple scales and resolutions. DoC operator nonlinearly removes redundancies and extracts fine grained details, texture of an image local structure and contrast to which HVS is highly sensitive. We introduce a new feature based on sparsity of DoC band. To extract perceptually important low-order structural information (edges), we use the non-oriented Difference of Gaussians (DoG) operator at different scales and resolutions. Measure of sparsity is calculated for DoG bands to get scalar features. To model the relationship between the extracted features and subjective scores, the general regression neural network (GRNN) is used. Quality predictions by the proposed DoC-DoG-GRNN model show higher compatibility with perceptual quality scores in comparison to the tested state-of-the-art metrics when evaluated on four benchmark datasets with synthesized views, IRCCyN/IVC image/video dataset, MCL-3D stereoscopic image dataset and IST image dataset.**

***Index Terms*—Difference of closings, DIBR synthesized view, granulometry, hat-transform, multi-resolution multi-scale image representation, quality prediction.**

I. Introduction

VIEW synthesis using depth-image-based rendering (DIBR) algorithms is important for 3D immersive imaging technologies such as multi-view and free-viewpoint video ith application in various fields: medicine, education, remote surveillance and entertainment. 3D video applications need a large number of views at different viewpoints. At the viewpoints, where video sequences captured by cameras are missing, new video sequences are synthesized from captured video and from their associated depth maps using DIBR algorithms. DIBR techniques introduce new types of distortions that are non-uniformly spread over the image, mainly in the disoccluded areas [1]. The disoccluded areas appear as black holes in image regions that become visible from the position of synthesized view, while invisible from the reference view position. Although the holes are filled synthetically by inpainting or interpolation technique, a texture distortion may appear. Beside the rendering process, distortions in synthesized images appear due to the distortions of texture images and depth maps, which may occur during the process of acquisition, compression and transmission. Distortions in the depth map may cause pixel projection to a wrong position and induce object shifting. The depth compression and view synthesis processes cause obvious distortions at object's edges. The distortions in texture cause intensity variations, while blurring and blocking distortions may appear all over the synthesized image.

The image quality assessment (IQA) of the synthesized view is of great importance for the development of view synthesis algorithms and 3D imaging applications. Traditional IQA metrics designed for natural scene images are unable to capture the non-uniform distortions introduced by synthesis algorithms [1]. The IQA models designed to evaluate distortions of synthesized images due to the rendering process only may not be effective in the evaluation of the complex distortions due to combined sources when texture images and depth maps are also distorted. In multi-view video systems, reference views are not available for all viewpoints. Therefore, referenceless quality metrics are highly desirable. In this paper, we propose a no-reference QA model which successfully evaluates distortions due to only a rendering process and complex distortions of synthesized images, which appeared due to the influence of the rendering process and the compromised quality of texture images and depth maps.

The Human Visual System (HVS) uses structural information from the viewing field for cognitive understanding. The disruption structure leads to the reduction of subjective image quality. The receptive fields of simple cells in mammalian primary visual cortex can be characterized as being selective to structure at different spatial scales, spatially localized and oriented [2]. Visual perception treats images on several levels of resolution simultaneously and this fact must be important in the study of perception [3]. By integrating information across spatial scales and resolutions, the HVS builds image representation. Local curvature measurements maybe performed by end-stopped neurons in the visual cortex [4]. The visual cortex is trying to produce an efficient representation in terms of extracting the statistically independent structure in images [5]. The goal of the primary visual cortex and other parts of the brain is to reduce redundancies of input visual signals in order to discover the intrinsic image structure, and thus create sparse image representation. Sparse image representations that capture the salient structures in line with the human perception are reported to be a processing strategy of the nervous system [6]. Sparsity is meant to extract "higher orders of statistical

dependencies", which cannot be simply described at the image pixel level [5]. We have taken into account the HVS characteristics in the design of the proposed IQA model.

To process the image and extract features from various scales, multi-scale approaches could be suitable. When the operator for generating image copies from fine to coarse scale is convolution by the Gaussian kernel, a linear scale-space can be built where the scale parameter is standard deviation of the Gaussian distribution [7]. By calculating the difference between the low-pass image copies of neighbor scales, the Difference of Gaussians (DoG) scale space of band-pass detail images is created. After convolution with the Gaussian kernel, the image is uniformly blurred, its edges are blurred and thus their localization may be imprecise. To preserve edges, unblurred and with precise localization, nonlinear morphological filters can be successfully used. Morphological filters are suitable for the construction of scale spaces since they remove structure from an image. The size of the structuring element is a scale parameter. Opening and closing multi-scale morphological filters have been used to create the scale-space of one-dimensional gray-scale signals [8]. Opening and closing are computationally simpler than the Gaussian filter [9]. Families of openings and closings, called granulometries, have been used in the morphological multi-scale approaches for shape-size distribution [10]. Using multi-scale opening and closing, some information is lost from one scale to the next: openings remove peaks of signals, while closings remove valleys. Peaks and valleys, which correspond to the critical points of a contour, are useful for visual perception. The residual of the closing compared to the original represents a bottom-hat transformation [11]. A sequence of hat transforms using structuring elements of increasing size has been used for the creation of the morphological curvature scale-space of 1D signals used for shape analysis [12]. Multi-scale hat transforms have been used for extraction of curvature extrema (as a representation for contour) through scales.

The multi-scale image representation at multiple resolutions using the DoG operator for extracting low-order structures has been used in the IQA model of 3D synthesized images [13]. In this paper, we introduce a novel QA model of DIBR-synthesized views which uses multi-scale image decomposition at multiple resolutions using both the DoG operator to extract low-order structures, and a newly proposed Difference of Closings (DoC) operator to extract higher-order structures (fine grained details, texture) of high curvature, which corresponds to the distortion of shapes to which the HVS is highly sensitive. DoC operator nonlinearly removes redundancies and extracts local structure and contrast to which HVS is highly sensitive. The proposed image decomposition using DoC operator can be described using granulometry by closing or using a multi-scale bottom-hat transform at multiple resolutions. The family of morphological closings with an array of line-shaped vertically-oriented structuring elements of increasing sizes is used to create a set of simplified signals by successively removing image structures (valleys) throughout scales. The size of a SE plays an important role in noise removal and geometrical detail preservation. We define the Difference of Closings (DoC) operator as the difference between morphologically smoothed images of neighboring scales. We believe that features extracted using both DoC and DoG operators at multiple scales and resolutions play complementary roles in characterizing view quality. The complex distortions of DIBR-synthesized images cause changes of both low and high-order structure at DoG and DoC bands, and their sparsities change. We measure sparsity of DoC and DoG bands through scales and resolutions using the Hoyer index [14] as a scalar feature. To model the relationship between the extracted features and subjective scores, a general regression neural network (GRNN) [15] is used. The trained network is then used to map extracted features to the quality score. The proposed DoC-DoG-GRNN model, tested on four datasets of 3D synthesized views, IRCCyN/IVC image dataset [1], MCL-3D stereoscopic image dataset [16], IST image dataset [69] and IRCCyN/IVC video dataset [17], shows high compatibility with perceptual quality scores, better than tested state-of-theart metrics.

In the next Section, we review metrics designed for quality assessment of DIBR-synthesized views which use multi-scale or/and multi-resolution image representation. In Section 3, the proposed model is described in more details. The datasets, evaluation criteria, model's parameters, overall performances and comparison to other metrics, as well as discussion and analysis, are presented in Section 4. The last Section concludes the paper.

II. Related Works

In this Section, we highlight some QA models designed for DIBR-synthesized images based on multi-resolution and/or multi-scale image representations in order to mimic the multi-resolution and multi-scale character of HVS. More complete overview of quality models designed for QA of DIBR-synthesized view is presented at [65].

A Laplacian type morphological pyramid with band-pass detail images with extracted edges has been used in the MP-PSNR metric [18], as well as its reduced version [19]. The difference between the appropriate detail images of the reference and the synthesized image pyramid at different resolutions is measured using the mean squared error to emphasize areas around edges that are prone to synthesis artifacts.

A Laplacian pyramid has been used for the extraction of luminance-based features to assess the naturalness of a 3D synthesized view in the newly proposed no-reference LVGC metric [20]. In order to capture local distortion, structure features extracted by second-order Gaussian derivatives and chromatic features are employed.

Low-pass images of a Gaussian type pyramid have been used for edge extraction in the Edge Intensity Similarity metric, EIS [21]. The similarity of edge intensity between appropriate edge images of the reference and the synthesized image pyramid through resolutions has been calculated. The Gaussian pyramid

of depth image has been used for the extraction of edge profiles in the no-reference QA metric proposed in [22]. Quality features have been extracted from the statistical distribution of edge regions. The first-order statistical features have been extracted in the gradient magnitude domain and the second-order statistical features have been extracted in the Laplacian-of-Gaussian domain. The extracted features have been used for building a random forest regression-based quality model. A multi-resolution low-pass image representation using Gaussian filters has been used in the reference-based LMS metric developed to evaluate the distortions in the whole synthesis process [23]. A low-level structural representation is calculated using the statistics of gradient intensity and orientation, while a mid-level structural representation is calculated using bag of words for contour description based on sparse coding. Gaussian filter has been applied for the generation of low-pass image copies at two scales for the extraction of features calculated from statistics of wavelet-based fusion of color and depth images in the QA model proposed to be applied before DIBR synthesis [58].

A multi-resolution image representation using downsampling has been used in the blind IQA model, designed using the philosophy that the DIBR-introduced geometry distortions damage the self-similarity characteristic of natural images and the damage degree tends to decrease as the resolution reduces [24]. In the following work, presenting a blind IQA model based on Multiscale Natural Scene Statistical analysis (MNSS) [25], it was found that the statistical regularity is destroyed in the DIBR-synthesized views. Estimating the deviation of degradations in main structures at different resolutions between a DIBR-synthesized image and the statistical model, the main structure damage can be quantified.

A morphological wavelet transformation has been used in the MW-PSNR metric, as well as its reduced version [26], [27]. Mean squared error has been used to measure the difference between appropriate subbands of the synthesized and the reference image. The high-high wavelet subband has been used in the computationally extremely efficient No-Reference Morphological Wavelet with a Threshold metric guided by the fact that DIBR-synthesized images are characterized by increased high frequency content [28].

Haar wavelet transformation has been performed on image blocks and image degradation has been measured using histograms of horizontal detail subbands in the reference-based IQA metric [29]. Discrete Wavelet Transform using CDF 9/7 filters has been used in the blind IQA model which quantifies geometric distortions, global sharpness, and image complexity [30].

A multi-scale image representation using the DoG operator at multiple resolutions has been used to extract a low-order structure, edges, and sparsity measure of DoG bands is calculated as a scalar feature in the general regression neural network-based DoG-GRNN model [13]. A multi-scale image DoG-representation has been used for extraction of two groups of features: the orientation selective statistics features and the texture naturalness features in the random forest regression based SET model designed to capture distortions in the whole view-synthesis process [31].

Multiscale representation has been used in the elastic metric and multi-scale trajectory based video quality metric which quantify the amount of temporal structure inconsistencies and unsmooth viewpoint transitions [66].

III. The Proposed QA Model

The proposed model is designed considering the properties of HVS: its high sensitivity to structure with a goal to reveal the intrinsic structure by reducing redundancies and creating sparse representation, but also its multi-resolution and multi-scale property when perceiving the visual scene. To extract a higher-order structure (fine grained details, texture) of high curvature, which corresponds to the distortion of shapes to which the HVS is highly sensitive, and to nonlinearly remove redundancies, we propose the morphological oriented Difference of Closings (DoC) operator and use it to create the sparse overcomplete multi-resolution and multiscale image representation (MR-MS-DoC) of non-negative integer valued DoC bands. We introduce the new feature based on sparsity of the DoC band. To extract a perceptually important low-order structure, edges, and to linearly reduce redundancies, we use non-oriented Difference of Gaussians (DoG) operator to create multi-resolution and multi-scale image representation (MR-MS-DoG). We believe that both low-order and high-order structure play complementary roles in characterizing the view quality. Since the DIBR-synthesized images are characterized by a structure-related distortion, their representations are with changed low-order and high-order structures and the sparsity of DoC bands and DoG bands are changed. The sparsity of DoC and DoG bands is measured using the Hoyer index as a scalar feature. The extracted features and subjective scores are used to train the general regression neural network (GRNN). The trained GRNN is then used to predict the quality score. The framework of the proposed model is shown on Fig. 1. Before explaining the proposed model in more details, we shortly review the basics of mathematical morphology used in model development.

A. Morphological Filters

Mathematical morphology is a powerful tool in image analysis mainly due to its nonlinearity and shape description properties. Morphological operations use a small pattern characterized by shape and size, called structuring element, to extract useful information. Morphological operators remove

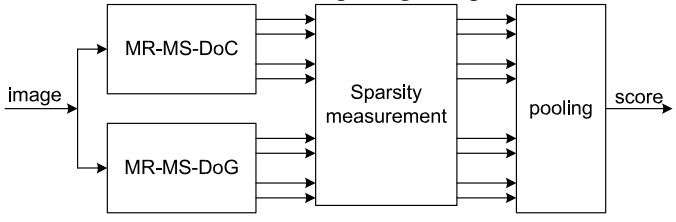


Fig. 1. The framework of the proposed DoC-DoG-GRNN model.

details from the signal without blurring the remaining structure. For grayscale image function $f(x, y)$ on Z^2, the basic morphological operators, dilation $f \oplus SE$(1) as expanding operation, and erosion $f \; SE$ (2) as shrinking operation, perform locally on the area defined by the structuring element SE using operators maximum and minimum [36].

$$(f \oplus SE)(x, y) = \max_{u,v \in SE} \{f(x - u, y - v)\} \quad (1)$$

$$(f \; SE)(x, y) = \min_{u,v \in SE} \{f(x + u, y + v)\} \quad (2)$$

The morphological filter closing C (3) locally modifies geometric signal features iteratively applying dilations and erosions, smooths signal eliminating specific image detail smaller than the structuring element without the global geometric distortion of unsuppressed features [36].

$$C_{SE} : (f \bullet SE)(x, y) = ((f \oplus SE) \; SE)(x, y) \quad (3)$$

Visually, the closing filter smooths contours of image dark regions, tends to fuse narrow breaks, eliminates small holes and fills gaps in the contours. Closing is invariant to translations of the structuring element and spatial invariance holds. The closing is extensive, $(f \bullet SE)(x, y) > f(x, y)$.

The residual produced after the application of closing filter to the image function $f(x, y)$ is Close-Top-Hat (CTH)(4), also called bottom-hat, or black-top-hat [11].

$$CTH : (f \bullet SE)(x, y) - f(x, y) \quad (4)$$

The CTH is region extraction operator which extracts dark details, valleys of the signal. The hat transformation provides an excellent tool for extracting features smaller than a given size from an uneven background.

B. Structuring Elements for Multi-Scale Morphology

For the creation of the morphological multi-scale image representation, an array of structuring elements of increasing size, SE_j, $1 \le j \le N$ is needed. The structuring element at scale j, SE_j, is created using the dilation of the structuring element at the first scale SE_1 with itself (5) [10]. The definition of scale corresponds to the spatial size of the structuring element.

$$SE_j = SE_1 \oplus \underbrace{SE_1 \ldots \oplus SE_1}_{j-1 \; times}, \; 2 \le j \le N \quad (5)$$

The shape of SE_j is determined by the shape of the primary pattern SE_1, and j controls the size.

We have explored the simplest case using an array of linear structuring elements with increasing length. The structuring element at the first scale is vertically oriented line pattern of length 2, Fig. 2. The structuring element length j corresponds to the scale $j - 1$. Line-shaped structuring elements used for the creation of multi-scale image representation allow the extraction of 1D line structures located in 2D image space.

Fig. 2. The primary pattern SE_1, line-shaped vertically oriented structuring element of length 2.

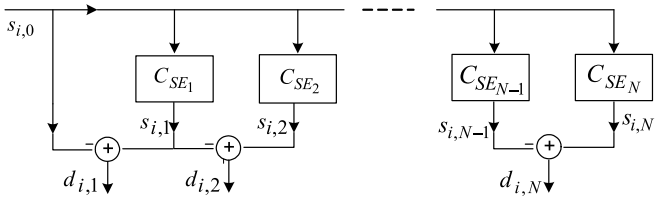


Fig. 3. The multi-scale image decomposition using morphological filter closing (C) with an array of linear structuring elements of increasing size SE_j, $j = 1,\ldots, N$,; $s_{i,0}$ the input image at resolution level i; $s_{i,j}$ the low-pass copy at scale j; $d_{i,j}$ the detail image (DoC band) at scale j.

C. Image Representation Using Granulometry by Closing

The granulometry is a useful tool for morphological multiscale image analysis and have been used for object size estimation, feature extraction for image segmentation and texture classification. A granulometry quantifies an amount of detail in an image at different scales.

We present the morphological multi-scale multi-resolution image decomposition to extract higher-order structure (fine grained details, texture) of high curvature which corresponds to the distortion of shapes to which the HVS is highly sensitive. The image decomposition can be described using the granulometry by closing at multiple resolutions. The multi-scale closings are useful smoothing filters because they preserve the shape and the location of vertical abrupt signal discontinuities (edges) [10]. At each resolution level i, $i = 1,\ldots M$, a series of low-pass image copies, $s_{i,j}$, $j = 1,\ldots, N$ (6) with decreasing details is created from the initial image, $s_{i,0}$, by applying a morphological filter closing, using an array of linear structuring elements with increasing length SE_j (5), as shown in Fig.3.

$$s_{i,j} = s_{i,0} \bullet SE_j, \qquad j = 1,\ldots, N \quad (6)$$

The morphological filter closing fills in the gulfs and the small holes of $s_{i,0}$ relative to SE_j. The size of the structuring element plays an important role in noise removal and geometrical detail preservation. With the scale increase, the length of the structuring element increases, more and more details are filtered out and each filtered image $s_{i,j}$ loose more and more structure. The monotonically increasing sequence of closings, $\{s_{i,j}\}$, $j = 1,\ldots, N$, constitutes a granulometry by closing. A granulometry by closing captures image structure darker than neighborhood. A fast algorithm has been proposed for efficient computation of linear granulometry [37].

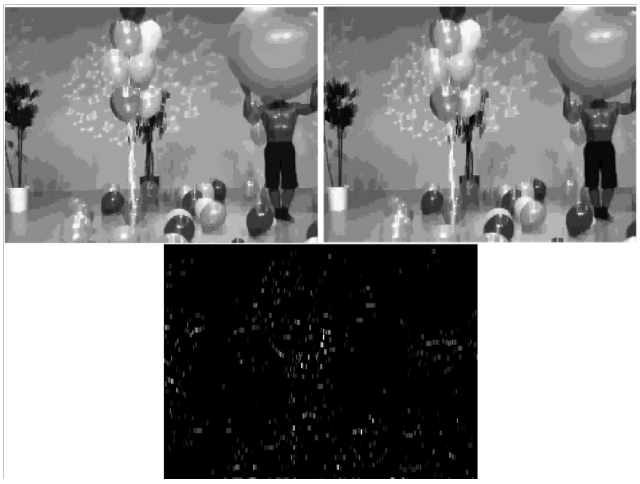


Fig. 4. The DoC band at resolution level 3 at scale 4, $d_{3,4}$, (bottom) of the image Balloon synthesized from the texture images and depth maps with JPEG distortion is created as the difference between smoothed images at scales 4 and scale 3 of resolution level 3, $d_{3,4} = s_{3,4} - s_{3,3}$: $s_{3,3}$ (first row left), $s_{3,4}$, (first row right).

We define Difference of Closings (DoC) operator (7) as the difference between morphologically filtered images of adjacent scales.

$$\text{DoC}: d_{i,j} = s_{i,j} - s_{i,j-1}, \qquad j = 1,..., N \tag{7}$$

The sequence of details removed between nearby closing filters, $\{d_{i,j}\}$, $j = 1,..., N$, creates a multi-scale image representation at resolution level i. Since the closing is an extensive operator, $s_{i,0} < s_{i,1} < \cdots < s_{i,N-1} < s_{i,N}$, the detail images, $d_{i,j}$, contain only non-negative integer values. The DoC band at resolution level 3 and scale 4 created from the image Balloon synthesized from the texture images and depth maps with JPEG distortion is shown on Fig. 4. The details extracted in DoC bands correspond to perceptually important building blocks of higher-order structure.

Morphological size distribution isolates features from noise at different scales. The difference of closings/openings at successive scale has been used in pattern spectrum as shape-size descriptor for multi-scale shape representation and description [10]. Morphological size distribution has been used for image noise reduction in the algorithm proposed for improved JPEG compression [38].

The multi-scale image transformation using DoC operator is reversible without information loss. The initial image at resolution level i, $s_{i,0}$, can be perfectly reconstructed from the detail images of all scales, $d_{i,j}$, $j = 1,..., N$ (all DoC bands) and the low-pass approximation at the highest scale, $s_{i,N}$ (8).

$$s_{i,0} = s_{i,N} - \sum_{j=1}^{N} d_{i,j} \tag{8}$$

To create the initial image for the next resolution level, $s_{i+1,0}$, the initial image from the current resolution level, $s_{i,0}$, is filtered using morphological filter closing with the structuring element SE_R and downsampled by a factor of two ($\sigma\downarrow$) by rows and by columns (9). The initial image at the first resolution level, $s_{1,0}$, is the image under test f. Initial images of all resolution levels, $s_{i,0}$, $i = 1,..., M + 1$, constitute the morphological pyramid of Gaussian type.

$$s_{1,0} = f$$

$$s_{i+1,0} = \sigma\downarrow(s_{i,0} \bullet SE_R)\ i = 1,..., M \tag{9}$$

Since the morphological operators used in multiresolution decomposition schemes involve only integers and only max, min and addition in their computation the calculation of morphological multiresolution decompositions have low computational complexity.

D. Image Representation Using Multi-Scale Hat- Transform

The close-top-hat (*CTH)* transform can be used to build the basic element of image semantic structure, and it is closely related to the perceptual quality of images. The hat transform can effectively capture image local structure and contrast to which the HVS is highly sensitive. The *CTH* operator enhances variations of pixel intensity in a neighborhood determined by the structuring element while preserving edges non-blurred. The hat transforms give important information about the fineness of variations of the gray level, with no need for image decomposition into harmonic frequencies. The hat transform has been used for extraction curvature, as one of the most powerful approaches for the representation and interpretation of objects in an image [12]. There is psychophysical, physiological and mathematical support in favor of using curvature as a representation for contours.

We can also describe MR-MS-DoC image decomposition using multi-scale hat transform at multiple resolutions. To create multi-scale close-top-hat transform at resolution level i,$\{cth_{i,j}\}$, $j = 1,..., N$, (10) we also use the granulometry by closing, $\{s_{i,j}\}$, $j = 1,..., N$ given in (6).

$$cth_{i,j} = s_{i,j} - s_{i,0}\, j = 1,..., N, \tag{10}$$

where the initial images at different resolutions, $s_{i,0}$, $i = 1,..., M+1$, are created as defined in (9). Since the closing is an extensive operator, the *CTH* transform creates detail image that contains pixels of non-negative integer values.

We can also define the Difference of Closings (DoC) operator as the difference between close-top-hat transforms of nearby scales (11), as shown in Fig. 5.

$$\text{DoC}: d_{i,j} = cth_{i,j} - cth_{i,j-1}\ j = 2,..., N$$

$$d_{i,1} = cth_{i,1} \tag{11}$$

where $d_{i,j}$ is the detail image (DoC band) at resolution level i and scale j. The sequences of details, $\{d_{i,j}\}$, at scales $j = 1,..., N$, and resolution levels $i = 1,..., M$ create multiresolution multi-scale DoC-based image representation.

E. Image Representation Using DoG Operator

Edges are important for visual perception playing a major role in the recognition of image content. The Difference of Gaussians (DoG) operator has been proposed as fast approximation of LoG operator used for edge extraction [39]. The DoG operator has been used in the DoG-GRNN model designed for the QA of the DIBR-synthesized images to extract low-order structure (edges) at multiple scales and resolutions [13].

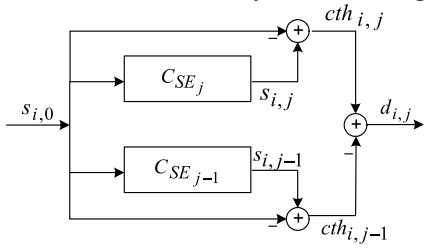


Fig. 5. The Difference of Closings (DoC) operator can be defined as the difference of close-top-hats of nearby scales, $di,j = cthi,j - cthi,j-1$: $cthi,j$ close-top-hat at resolution level i and scale j, di,j the detail image (DoC band) at resolution level i and scale j, $si,0$ the input image at resolution level i.

In this work, DoG-based image representation of low-order structure (edges) is used, together with the DoC-based image representation of higher-order structure (fine details, texture) to improve the IQA model's performances. At each resolution level i, $i = 1,... P$, an array of Gaussian smoothed images, $g_{i,j}$, $j = 1,..., Q$ is generated by convolution of the input image $g_{i,0}$ with the Gaussian function $G(x, y,\sigma_j)$ using an array of standard deviations of the Gaussian distribution σ_j, $j = 1,..., Q$ as scale parameters (12).

$$g_{i,j}(x, y) = {}_{i,0}(x, y) * G(x, y,\sigma_j), \qquad j = 1,..., Q$$

$$G(x, y,\sigma_j) = \frac{1}{2\pi\sigma^2 e^{j}} \qquad -\frac{x}{2\sigma^2 2_{\pm}y2} \tag{12}$$

The Difference of Gaussians (DoG) operator is defined as the difference of two nearby Gaussian smoothed images (13).

$$\text{DoG}: e_{i,j} = g_{i,j-1} - g_{i,j} \tag{13}$$

The DoG band $e_{i,j}$, is a band-pass filtered image which contains all frequencies between the cut-off frequencies of the two Gaussians which correspond to edge lines. Edges of different fineness are detected at different scales as well as the noise due to synthesis distortion. The DoG bands at resolution levels 1 and 5, at scale 2, created from the image Balloon synthesized from the texture images and depth maps with JPEG distortion are shown on Fig. 6. DoG bands of higher resolution contain enhanced edges while DoG bands of lower resolution remind saliency maps.

The initial image for the next resolution level, $g_{i+1,0}$ is generated from the last Gaussian smoothed image of the current resolution level $g_{i,Q}$ by downsampling it with a factor of two by rows and by columns (14).

$$g_{1,0} = f$$

$$g_{i+1,0} = \sigma\downarrow(g_{i,Q}), \qquad i = 1,..., P - 1 \tag{14}$$

The multi-resolution multi-scale DoG-based image representation consists of detail images $e_{i,j}$ at resolution levels i, $i = 1,... P$, and scales j, $j = 1,..., Q$.

F. Measure of Sparsity

Sparsity of signal representation has been important concept in signal analysis, compression, sampling used in diverse areas such as image processing, medical imaging and face recognition. The Hoyer index, calculated as normalized relationship between l_1 and l_2 norms, has been proposed as a measure of sparseness of the image representation obtained using non-negative matrix factorization [14]. The Hoyer index quantifies how much energy is packed into only a few components. It takes value 0 when all coefficients are equal and it is 1 when only a single coefficient is non-zero. The It satisfies five heuristic criteria (out of six) used for comparing sparseness measures [40]. Hoyer index has been used in the IQA models of DIBR-synthesized images based on morphological pyramid, morphological wavelets [41] and DoG-based image representation [13]. We use the Hoyer index to measure sparsity of DoC and DoG bands in the proposed model. The Hoyer indices of DoC and DoG bands from the interval [0, 1] are concatenated and form the neural network input vector to GRNN.

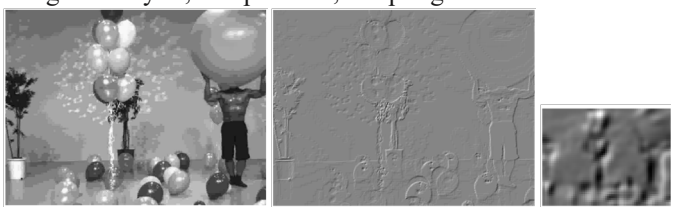


Fig. 6. The DoG band at scale 2 of resolution level 1, $e1,2$ (in the middle) and at scale 2 of resolution level 5, $e5,2$ (right), created from the image Balloon synthesized from JPEG distorted content (left).

G. Feature Pooling

In the field of IQA, machine learning techniques are widely used to model the human perception achieving excellent performance as they build functional relationship between the image features and subjective quality scores. A Generative Adversarial Network has been used in the no-reference quality metric of free-viewpoint images which uses masks to mimic dis-occluded regions [67]. A General Regression Neural Network (GRNN) as a powerful regression tool based on statistical principles shows good results even with smaller learning datasets and is particularly advantageous with sparse data in real-time environment [15]. GRNN has been used to model the relationship between perceptually relevant features and subjective scores in IQA model which handle multiple distortion [81]. It has been used in the IQA models of DIBR-synthesized images [41], [13]. GRNN has been compared to support vector regression algorithm (SVR) in the models proposed for IQA of synthesized images based on morphological pyramid decomposition and morphological wavelet [80]. It has been shown that SVR-based model shows lower performances than GRNN-based model. We have used GRNN to model the relationship between the extracted features and subjective scores. The learned model is then used to predict the image/video quality. The only parameter is the spread, that represents the width of radial basis function of GRNN. The GRNN has been implemented using Matlab function *newgrnn.*

IV. Experimental Results

In this section, used datasets and evaluation criteria are shortly described. We've described the choise of model's parameters and show the performances of the proposed model calculated using datasets. We've also presented the parameter set of the proposed model for the prediction of images synthesized from compressed content with improved computational efficiency. The proposed DoC-DoG-GRNN model is compared to other IQA models and their performances are analyzed.

A. Datasets and Evaluation Criteria

The proposed metric is evaluated using publicly available datasets of DIBR-synthesized images and videos, namely the IRCCyN/IVC DIBR image dataset [1], [43], the MCL-3D stereoscopic image dataset [16], [44] and the IRCCyN/IVC DIBR video dataset [17], [45].

IRCCyN/IVC image/video dataset contains 12 reference frames/videos from three multi-view video plus depth sequences and 84 frames/videos synthesized using seven DIBR-synthesis algorithms. The frames are synthesized by a single-view synthesis (a frame at the "virtual viewpoint" is synthesized using captured texture frame and its associated depth frame from single viewpoint). The images/videos of these two datasets are synthesized from undistorted data, so the focus is on rendering artifacts. Video sequences BookArrival is of 15 frames/s rate while Lovebird and Newspaper are of 30 frames/s rate. The frames resolution is 1024 × 768 pixels. Mean Opinion Scores (MOS) are provided for both the synthesized and for the reference frames/videos. The Difference Mean Opinion Scores (DMOS) is calculated by measuring the difference between the reference and the synthesized image MOS.

The MCL-3D dataset contains 684 synthesized stereoscopic image pairs synthesized from nine image-plus-depth sources and associated mean opinion score (MOS) values. One third of images are of resolution 1024 × 728 and two thirds are of resolution 1920 × 1080. The majority of images (648 stereo images) are synthesized using both left and right views distorted by one of six distortion types: Gaussian blur, additive white noise, down-sampling blur, JPEG and JPEG-2000 compression and transmission error. Distortions have been symmetrically applied to left and right viewpoints to either the texture images or the depth maps or both texture and depth maps at four distortion levels before stereoscopic image rendering using the View Synthesis Reference Software [74]. The minority of images (36 stereo images) have been synthesized from undistorted texture and depth maps of single view using four DIBR synthesis algorithms. They contain only the distortions caused by imperfect rendering.

In order to evaluate performances of the proposed model and other metrics, three evaluation criteria were used: Root Mean Squared Error (RMSE) to compute the prediction error, Pearson Linear Correlation Coefficient (PLCC) to compute prediction accuracy and Spearman Rank-order Correlation Coefficient (SROCC) for prediction monotonicity. The performances of the proposed model and all tested metrics are calculated using the Differential Mean Opinion Score (DMOS) for the IRCCyN/IVC image/video dataset and using the Mean Opinion Score (MOS) for the MCL-3D dataset. A novel methodology for performance evaluation which takes into account statistical significance of subjective scores, with the ability to analyse the data statistically after merging results from multiple datasets, can be used when the standard deviation of subjective scores is provided [68].

B. Performance Calculation Using Train-Test Process

In the evaluation of the proposed model, the k-fold cross validation strategy is used to split the dataset with D images to k disjoint test subsets of similar size, each with D/k images (for k = 5 it is 20% images of the dataset), and to select the train subsets with D-D/k images (for k = 5 it is 80% images of the dataset) such that there is not overlap between the test and the train subsets. GRNN has been trained by mapping the features calculated from the train subset to the subjective scores. Then, the trained model has been used to predict the quality of the images from the test subset. The k-fold cross validation procedure is performed 1000 times to avoid over-fitting. We've analyzed the 5 ways for the performance calculation marked as Case1, Case1A, Case 2, Case 2A and Case 2B.

Case 1: At each iteration of the k-fold cross validation strategy, GRNN is used to predict the scores of the test subset. Then, the performances of the test subset predicted scores are

calculated. The median value of the performances through cross-validation iterations and through 1000 repetitions is calculated as the final model's performances.

Case 1A: The only difference to Case 1 is that the test subset predicted scores are additionally nonlinearly mapped to the predicted $DMOS_p$ using a five-parameter function (15) before performance calculation.

$$DMOS_p = a \cdot score^4 + b \cdot score^3 + c \cdot score^2 + d \cdot score + e \quad (15)$$

The parameters *a,b,c,d,e* are obtained through regression to minimize the difference between *DMOS* and *DMOSp*.

Case 2: At each iteration of the k-fold cross-validation strategy, GRNN is used to predict the scores of the test subset. Then, we concatenate the predicted scores from k iterations of single cross-validation process to get the array of the whole dataset predicted scores. We calculate the median of the whole dataset predicted scores through 1000 repetitions. Finally, we calculate the performances on the median of the whole dataset predicted scores as the final model's performances.

Case 2A: The only difference to Case 2 is that median of the whole dataset predicted scores are additionally nonlinearly mapped using a function (15) before performance calculation.

Case 2B: The only difference to Case 2 is that the test subset predicted scores are additionally nonlinearly mapped using a function (15) before concatenation to the whole dataset scores.

C. Model's Parameters and Performances

In this section, parameters selection is described. Parameters of the DoC-DoG-GRNN model are the number of resolution levels and the number of scales of DoC-based and DoG-based decompositions and the spread that represents the width of radial basis function of GRNN. For more efficient calculation, we've first explored the parameters of DoC-based decomposition using the DoC-GRNN model. Then, using the selected parameters of DoC-based decomposition, we've explored the parameters of DoG-based decomposition using DoC-DoG-GRNN model. The selected parameters allow the highest model's performances. We have selected parameter set 1 using IRCCyN/IVC image dataset. Then, using the selected parameters, we've calculated the performances of the proposed model for three datasets.

We've explored performances of the DoC-GRNN model using different number of resolution levels (1-7), scales (1-7) and different values of spread parameter (0.001-0.05). The best performances are achieved when 5 resolution levels and 5 scales are used. Performances are further improved when the DoC bands from the first scale, $d_{i,1}$, and the second scale, $d_{i,2}$, are omitted, and three DoC bands at scales 3-5, $d_{i,3} - d_{i,5}(7)$, i = 1,...,5, of each resolution level are used. To create initial images for multi-scale image decomposition at different resolutions, morphological close filter with lineshaped vertically oriented structuring element of length 2, $SE_R = SE_1$, is used. The sparsity of 15 DoC bands is calculated using the Hoyer index as a scalar feature. The lowpass image copy of the lowest resolution level, $s_{6,0}$, is also used for the feature extraction and the total number of features of DoC-GRNN model is 16. In order to improve model's performances, we've also used image representation generated using the DoG operator to extract low-order structure at multiple scales and resolutions. For the generation of DoG bands at scales j, j = 1,..., Q, Gaussian function using an array of standard deviations, σ_j, is applied. Standard deviations at different scales, σ_j, are calculated such that $\sigma_j = k^{j-1}$, where constant multiplicative factor k is determined by the number of low-pass images at one resolution level, Q: $k = 2^{1/Q}$. The selected size of the Gaussian kernel window is integer$(6 * \sigma_j)$. We have explored the performances of the DoC-DoG-GRNN model using DoC-based decomposition with 5 resolution levels and scales 3-5 at each level and DoG-based decomposition with different number of resolution levels (1-7) and scales (1-7), with different values of spread parameter (0.001-0.05). Although high performances of the combined DoC-DoG-GRNN model, with PLCC higher than 0.88, is achieved using different number of resolution levels and scales of DoG-based decomposition, we've selected P = 5 resolution levels and Q = 6 scales to achieve the highest value of SROCC. From DoG bands, $e_{i,1}-e_{i,6}$, i = 1,...,5, 30 scalar features are calculated using Hoyer index as a measure of sparsity.

The GRNN input vector of the combined DoC-DoG-GRNN model is created by concatenation of 16 features calculated from DoC-based decomposition and 30 features calculated from DoG bands. The SROCC of the combined DoC-DoGGRNN model, DoC-GRNN and DoG-GRNN models with fluctuation of the GRNN spread parameter for three datasets, IRCCyN/IVC image/video and MCL-3D dataset, are shown on Fig. 7. High performances of DoC-GRNN model are achieved mostly thanks to the capacity of morphological series to capture higher order properties of spatial random processes. The combined model, DoC-DoG-GRNN, shows better performances than the DoC-GRNN and DoG-GRNN models, for all datasets. From the results we conclude that both low-order structure and high-order structure extracted

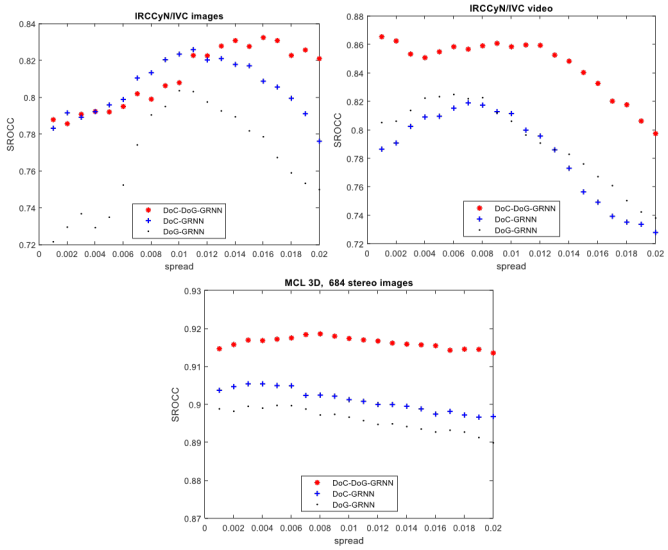


Fig. 7. SROCC of the combined DoG-DoC-GRNN model, DoC-GRNN and DoG-GRNN models for IVC datasets (up) and MCL-3D dataset (down).

TABLE I

PERFORMANCES OF THE PROPOSED DoC-DoG-GRNN MODEL CALCULATED ON 5 WAYS

IRCCyN/IVC image			
	PLCC	SROCC	RMSE
Case 1	0.8805	0.8309	0.3298
Case 1A	0.9107	0.8357	0.2599
Case 2	0.8733	0.8285	0.3300
Case 2A	0.8786	0.8323	0.3179
Case 2B	**0.9334**	**0.8891**	**0.2402**
MCL-3D			
	PLCC	SROCC	RMSE
Case 1	0.9242	0.9159	1.0146
Case 1A	0.9274	0.9159	0.9700
Case 2	0.9442	0.9365	0.8686
Case 2A	0.9450	0.9365	0.8509
Case 2B	**0.9491**	**0.9402**	**0.8221**
IRCCyN/IVC video			
	PLCC	SROCC	RMSE
Case 1	0.8552	0.8485	0.3581
Case 1A	0.8927	0.8498	0.2884
Case 2	0.8370	0.8395	0.3659
Case 2A	0.8423	0.8400	0.3567
Case 2B	**0.9130**	**0.9035**	**0.2744**

at multiple scales and resolutions should be considered for improved results in quality assessment of DIBR-synthesized views. We've selected the spread parameter 0.014 as the value for the best performances of IVC image dataset.

Performances of the proposed DoC-DoG-GRNN model for three datasets, IRCCyN/IVC image/video and MCL-3D dataset, calculated on five ways, Case 1, Case 1A, Case 2, Case 2A and Case 2B are shown in Table I. The best performances for all datasets are achieved in Case 2B.

D. The Sensitivity on the Cross-Validation Partition Ratio

We've tested the performances of the proposed QA model using three partition ratio, k = 10, 5, 4 of the k-fold

TABLE II

MODEL'S PERFORMANCES FOR DIFFERENT DATA PARTITION RATIOS

IRCCyN/IVC image				
Cross-validation ratio	k	PLCC	SROCC	RMSE
90% - 10%	10	0.9019	0.8441	0.3021
80% - 20%	5	0.8805	0.8309	0.3298
75% - 25%	4	0.8724	0.8218	0.3364
MCL-3D				
Cross-validation ratio	k	PLCC	SROCC	RMSE
90% - 10%	10	0.9368	0.9240	0.9307
80% - 20%	5	0.9242	0.9159	1.0146
75% - 25%	4	0.9176	0.9103	1.0566
IRCCyN/IVC video				
Cross-validation ratio	k	PLCC	SROCC	RMSE
90% - 10%	10	0.8903	0.8609	0.2871
80% - 20%	5	0.8552	0.8485	0.3581
75% - 25%	4	0.8424	0.8381	0.3673

cross-validation strategy, and the results calculated according to the Case 1 procedure are shown in Table II. With decrease of the training data partition size, the prediction performances decrease, although the partition ratio, k, has small influence on performance. For all tested partitions, SROCC is higher than 91% for MCL-3D dataset and higher than 82% for IRCCyN/IVC image/video dataset. Performances shown in all other tables are calculated using k = 5.

E. Comparison of the Proposed Model to Other IQA Models

To validate the effectiveness of the proposed model, it is compared to the models designed to evaluate the quality of DIBR-synthesized images and to the models designed to evaluate the quality of natural images. In the evaluation of the training-free models, their scores are nonlinearly mapped to the predicted scores using a cubic logistic function [46]. In the calculation of video sequence score using training-free metrics, median time pooling of frame's scores is applied, using all frames. Performances of the proposed and other tested models evaluated on IRCCyN/IVC image/video dataset and MCL-3D stereo image dataset are presented in Table III. Table IV contains performances of the metrics designed for QA of DIBR-synthesized images with rendering distortions only, taken from the papers where they were proposed.

The first group of metrics used for comparison designed to evaluate quality of DIBR-synthesized images, which use machine learning-based techniques, contains no-reference

models, LVGC [20], SET [31], CSC-NRM [42] and referencebased models, Q [33], Q [32], ST-IQM [34] and LMS [23]. The proposed DoC-DoG-GRNN model is a no-reference machine learning-based technique. In the evaluation of the DoC-DoG-GRNN model, cross-validation process described in III B is applied and the performances are calculated as it is described in Case 2B. In the calculation of video sequence score, we've used non-linear time pooling of frame's features. The best performances of the proposed DoC-DoGGRNN model are achieved when median time pooling is used for DoC-based features and max time pooling is used for DoG-based features. We've calculated the features of all frames. The performances of the random-forest-based metrics designed to evaluate quality of DIBR-synthesized images with distortions due to whole synthesis process, LVGC using 310 features and SET using 51 features, are taken from [20] and [31]. The Convolutional Sparse Coding-based No-Reference Model, CSC-NRM (shown in Table IV), uses convolutional sparse coding to learn from local regions and computes a sparse representation of an image [42].

The performances of the reference-based models designed to evaluate the quality of DIBR-synthesized images using machine learning techniques, Q [33], Q [32], ST-SIQM [34] and LMS [23] are taken from the papers where they were introduced. Q [33] quantifies the impact of structure-related distortion on perceived quality of synthesized images using the low-level contour descriptor, the mid-level contour category descriptor and the task-oriented non-natural structure descriptor. Q [32] estimates the degree of texture and geometric distortions using coarse-scale and fine-scale features using the fact that the HVS first produces a coarse perception of the global image, and then focuses on specific local areas for a fine perception of image quality. Sketch Token-based Synthesized Image Quality Model, ST-SIQM [34], measures how classes of contours change after synthesis from higher semantic level using a mid-level contour descriptor.

The metrics from the first group can reliably predict the quality of DIBR-synthesized images from IRCCyN/IVC image dataset. The proposed model DoC-DoG-GRNN outperforms all other methods from all groups, achieving high performances, PLCC = 0.93, SROCC = 0.889, in the evaluation of IRCCyN/IVC image dataset. The metrics DoC-DoG-GRNN, LVGC, SET, Q [32] and LMS [23] can also reliably evaluate the MCL-3D dataset. DoC-DoG-GRNN, LVGC and SET models achieve high performances in the evaluation of MCL-3D dataset. The proposed DoC-DoG-GRNN model is comparable to the best state of the art LVGC model in the evaluation of MCL-3D datasets. The proposed DoC-DoG-GRNN model can also evaluate the IRCCyN/IVC video dataset with high performances, PLCC = 0.9130, SROCC = 0.9035, outperforming all other methods from all groups.

The second group of metrics contains training-free metrics designed to evaluate DIBR-synthesized images with rendering distortions, namely blind metrics: Q [30], MNSS [25], Q [60], OMIQA [47], APT [48], NIQSV [49], NIQSV+ [50] and reference-based metrics Q [59], LoGS [35], PSNR' [64], EIS [21], EM-IQA [61], MP-PSNR [19], CT-IQA [62], MW-PSNR [26], BF-M [63]. The performances of MNSS are taken from [25]. The joint blind quality assessment and enhancement algorithm for 3D synthesized images Q [60] use a global predictor, Kernel Ridge Regression, to identify the surface of the regions with geometric distortions with low prediction error and integrates the Natural Image Quality Evaluator to judge the structural distortions. Efficient OMIQA [47] capture the geometric and structural distortions by identifying and removing geometrically distorted pixels, outliers, using non-linear median filtering. Autoregression Plus Threshold (APT) model [48] captures the geometry distortion by calculating the error between the DIBR-synthesized image and its auto-regression-based local prediction, and uses the threshold to highlight the most important regions. The five parameter nonlinear logistic function has been applied to APT scores before performances calculation as in [48]. NIQSV [49] and NIQSV+ [50] calculate the error between the DIBR-synthesized image and its morphologically filtered approximation and use morphological gradient to highlight the most important edge regions. NIQSV+ additionally detects black holes and stretching distortion. NIQSV+ is calculated with the following parameters: color weight K1 = 0.5, black hole weight k_z = 200 and C = 1.

Reference-based Q [59] captures holes and strip distortions to characterize the local quality of DIBR-synthesized image by analyzing local similarity and estimates the global sharpness using the Just Noticed Difference. LoGS [35] quantifies local

TABLE III

PERFORMANCES OF THE PROPOSED DOC-DOG-GRNN MODEL AND OTHER METRICS

		IRCCyN/IVC (84 images)			**MCL-3D (684 stereo images)**			**IRCCyN/IVC (84 videos)**		
		PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
no-reference IQA models with training for DIBR images	**DoC-DoG-GRNN**	**0.9334**	**0.8891**	**0.2402**	**0.9491**	0.9402	**0.8221**	**0.9130**	**0.9035**	**0.2744**
	LVGC [20]	0.8606	0.8237	0.3118	0.9453	**0.9426**	0.8578	-	-	-
	SET [31]	0.8586	0.8109	0.3015	0.9117	0.9108	1.0631	-	-	-
reference-based IQA with training for DIBR images	Q [32]	0.8714	0.8453	0.2790	0.8390	0.8330	1.4049	-	-	-
	LMS [23]	0.807	0.776	0.391	0.837	0.833	1.476	-	-	-
blind IQA models for DIBR images	MNSS [25]	0,785	0.770	0.412	-	-	-	0.632	0.631	0.382
	OMIQA [47]	0.7379	0.7054	0.4494	0.4214	0.3337	2.3663	0.6845	0.6411	0.4824
	APT [48]	0.7354	0.7147	0.4512	0.3806	0.1827	2.4059	0.6520	0.6287	0.5017
	NIQSV+ [50]	0.7165	0.7016	0.4645	0.3259	0.3354	2.4668	0.5671	0.5137	0.5450
	NIQSV [49]	0.6297	0.5866	0.5172	0.6790	0.6428	1.9157	0.6188	0.5350	0.5198
reference-based training-free IQA models for DIBR images	Q [59]	0.8366	0.8147	0.1341	0.7872	0.7445	1.5965	-	-	-
	LoGS [35]	0.8256	0.7812	0.3601	0.7612	0.7577	1.6874	-	-	-
	EIS [21]	0.7288	0.7259	0.4559	-	-	-	0.5871	0.6174	0.4562
	MP-PSNR [19]	0.6954	0.6629	0.4784	0.7742	0.7919	1.6514	0.5573	0.5513	0.5494
	MW-PSNR [26]	0.6625	0.6232	0.4987	0.7679	0.7853	1.6713	0.5464	0.4957	0.5542
reference-based training-free IQA models for natural images	PSNR	0.4557	0.4417	0.5927	0.7779	0.7900	1.6397	0.4781	0.4324	0.5812
	SSIM	0.4292	0.3929	0.6014	0.7539	0.7741	1.7144	0.6123	0.5902	0.5231
	MS-SSIM [51]	0.5411	0.4833	0.5599	0.8386	0.8518	1.4213	0.6482	0.5930	0.5038
	IW-SSIM [52]	0.5011	0.4513	0.5762	0.8694	0.8878	1.2894	0.6188	0.5455	0.5198
	GSM [53]	0.4315	0.3858	0.6006	0.8482	0.8603	1.3822	0.5311	0.4927	0.5607
	GMSD [54]	0.4890	0.3858	0.5808	0.8413	0.8472	1.4103	0.5233	0.4943	0.5638
	PSIM [55]	0.4049	0.4208	0.6088	0.8693	0.8798	1.2898	0.5490	0.4881	0.5530
	VSI [56]	0.4638	0.4084	0.5899	0.8717	0.8855	1.2786	0.5633	0.5227	0.5467
	SSRM [57]	0.4777	0.4005	0.5849	0.8790	0.8962	1.2442	0.5256	0.5192	0.5629

geometric distortions in disoccluded regions and global sharpness has been quantified using a reblurring-based strategy. The simple weighted PSNR, PSNR' [64], compensates the object shift and uses a disparity map as a mask to weight the final distortion. The performances of multi-scale edge intensity similarity metric, EIS, are taken from [21].

TABLE IV

PERFORMANCES OF METRICS DESIGNED FOR DIBR-SYNTHESIZED IMAGES

IRCCyN/IVC (84 images)				
	metric	PLCC	SROCC	RMSE
no-reference with training	**DoC-DoG-GRNN**	**0.9334**	**0.8891**	**0.2402**
	CSC-NRM 42	0.8302	0.7827	0.3233
reference-based with training	Q 33	0.9023	0.8448	0.2870
	ST-SIQM 34	0.8217	0.7710	0.3929
blind IQA models	Q 30	0.7995	0.7867	0.4000
	Q 60	0.8054	0.7598	0.3946
reference-based training-free	PSNR' 64	0.8242	0.7889	0.3771
	EM-IQA 61	0.7430	0.6626	0.4455
	CT-IQA 62	0.6809	0.6626	0.4877
	BF-M 63	0.6980	0.5885	0.4768

Elastic metric, EM-IQA [61], quantifies the deformations of curves in the local distortion regions. Reduced version of MP-PSNR [19] is calculated using only detail images from resolution levels 4 and 5 generated using structuring element of size 5 × 5 and the low-pass copy of the lowest resolution. A context tree based metric CT-IQA [62] measures how the structure change due to synthesized artifacts by measuring the dissimilarity in encoding cost between the original and synthesized image. Reduced version of MW-PSNR [26] uses only 11 wavelet sub-bands from resolution levels 4-7. Before wavelet decomposition, larger images of MCL3D dataset of size 1088 × 1920 are resized to 1024 × 1920. A bilateral filtering based model BF-M [63] use a contour, a shape and a texture based estimators to quantify the amount of structural and textural change in synthesized image.

The training-free metrics of the second group can successfully predict quality of images from IRCCyN/IVC image dataset although there is a need for the performance improvement. Blind metrics Q [30], Q [60], MNSS and reference-based metrics, Q [59], LoGS and PSNR' show higher performances when evaluating IRCCyN/IVC image dataset. Blind models OMIQA, APT and NIQSV+ show low performances when they evaluate MCL-3D dataset. Reference-based metrics, Q [59], LoGS, MP-PSNR and MW-PSNR, can be used for the evaluation of MCL-3D dataset but with limited performances. Metrics from this group show lower performances in prediction of the video quality. Among training-free metrics, OMIQA [47] achieves the best performances in the evaluation of IRCCyN/IVC video dataset.

The third group of metrics contains reference-based metrics designed for the evaluation of natural images. We've tested classic PSNR, structure-based metrics, SSIM, MS-SSIM [51], IW-SSIM [52], GSM [53], GMSD [54], PSIM [55], saliency-based metric VSI [56] and sparseness-based SSRM model [57]. MS-SSIM [51] combines similarity measures calculated at different resolutions of low-pass pyramids. Information content Weighting SSIM measure, IW-SSIM [52], combines weighting maps of information content extracted from a Laplacian pyramid and multi-scale structural simularity measure.

The Gradient Structure Metric (GSM) [53] uses gradient similarity to measure the change in contrast and structure. The Gradient Magnitude Similarity Deviation (GMSD) model [54] predicts image quality using gradient magnitude similarity (GMS) between the reference and the distorted images combined with the pooling using standard deviation of GMS map. The Perceptual SIMilarity (PSIM) model [55] uses the gradient magnitude maps at two scales, at a small scale to measure the variations in macrostructure in order to emulate the human's basic perception, and at a large scale to reflect the variations in microstructure and to emulate the human's detailed perception. Visual Saliency-based Index (VSI) [56] use visual saliency map (it can reflect how "salient" a local region is to the HVS) as a feature map to characterize the quality of local image regions and as a weighting function at the score-pooling stage. Sparseness Significance Ranking Measure model (SSRM) [57] is based on the sparsity hypothesis that the visual quality assessment should be compatible with sparse coding, which is one of the main strategies of redundancy reduction implemented in our brain.

TABLE V

CROSS-VALIDATION OF THE PROPOSED MODEL

training dataset	test set	PLCC	SROCC	RMSE
IRCCyN / IVC (84 images)	IVY subset1 (21 images VSRS 76)	0.7894	0.6948	8.3085
	IVY subset2 (21 images alg 77)	0.7175	0.6312	10.4457
	IVY subset3 (21 images alg 78)	0.6785	0.6364	10.8465
	IVY subset4 (21 images alg 79)	0.5708	0.5610	11.3490
IVY dataset, (84 images)		0.7288	0.6789	9.4332

The metrics of the third group can be successfully used to assess the quality of images from MCL-3D dataset. Among these models, SSRM shows the best performances in the evaluation of MCL-3D dataset. The metrics of this group show low performances when they are used to predict the quality of images/videos from the IRCCyN/IVC image/video dataset.

F. Generalization Ability

We've explored the generalization ability of the proposed DoC-DoG-GRNN model using the IRCCyN/IVC dataset for training and subsets of IVY dataset [75] for testing. IVY dataset contains 84 stereo images synthesized from undistorted content of 7 sources using 4 DIBR synthesis algorithms [76]–[79] by view extrapolation using only one reference view. IVY dataset has been created to study the effects of binocular asymmetry caused by mismatch between left and right image on quality of synthesized stereo images.

Four subsets of IVY dataset are created considering four DIBR synthesis algorithm. The best performances are achieved in evaluation of subset 1, which contains images synthesized using VSRS algorithm [76], as shown in Table V. The lowest performances are achieved in evaluation of subset 4, which contains images synthesized using algorithm [79]. Algorithm [79] achieves consistency between the virtual views without distortion of depth values and allows generation of high quality image. The proposed model fails in the evaluation of synthesized images of high quality (without shapes and edges distortion). The performances of the proposed model evaluated on IVY dataset, when it is used for both training and testing calculated as case1A, are also shown in the Table V.

G. QA of Images Synthesized From Distorted Content

In Section IV C, we've selected model's parameters (parameter set 1) using IRCCyN/IVC image dataset which contains images synthesized from undistorted content. It has been shown that the proposed model using parameter set 1 shows high performances when predicting the quality of images of MCL-3D dataset, mainly synthesized from distorted content. For MCL-3D dataset, slightly better performances are achieved when the proposed model uses parameter set 2: DoC-based decomposition at 5 resolution levels with scales 2-7 at each resolution level, $d_{i,2}$-$d_{i,7}$, ^{i} = 1,...5, DoG-based decomposition at 5 resolution levels with scales 3-6 at each resolution

TABLE VI

PERFORMANCES OF DOC-DOG-GRNN MODEL USING PARAMETER SET 2 FOR MCL-3D SUBSETS BY MVD SOURCES

size	# of images	MVD source	PLCC	SROCC	RMSE
1024x768	76	Kendo	0.9633	0.9294	0.7474
1920x1088	76	GT_fly	0.9564	0.9255	0.7927
1920x1088	76	PoznanStreet	0.9509	0.9249	0.8451
1024x768	76	Baloons	0.9497	0.9184	0.8557
1920x1088	76	MicroWorld	0.9444	0.9312	0.9009
1024x768	76	Love_bird1	0.9356	0.9071	1.0130
1920x1088	76	Dancer	0.9333	0.9241	1.0296
1920x1088	76	Poznan_hall2	0.9237	0.8876	1.0521
1920x1088	76	Shark	0.8950	0.8857	1.2111

TABLE VII

MODEL'S PERFORMANCES FOR MCL-3D SUBSETS WITH DIFFERENT DISTORTION TYPES

distortion type	# of images	PLCC	SROCC	RMSE
downsampling blur	108	0.9586	0.9467	0.8803
blur	108	0.9556	0.9331	0.8288
JPEG2000	108	0.9536	0.9081	0.7758
JPEG	108	0.9352	0.8537	0.8329
additive white noise	108	0.8656	0.8454	1.3000
transmiss. errors	108	0.8203	0.7957	1.2181

level, $e_{i,3}$-$e_{i,6}$ ^{i} = 1,...,5, and spread 0.09. The number of features using parameter set 2 is $5*6 + 1 + 5*4 = 51$.

We've analyzed performances of the proposed DoC-DoGGRNN model using parameter set 2 by nine MVD sources with images of smaller size, 1024 × 768, Baloon, Kendo, Love_bird1, and with images of larger size, 1920 × 1088, Shark, GT_fly, MicroWorld, Dancer, PoznanStreet, Poznan_hall2, each subset with 76 stereo images. The performances slightly vary by sources as shown in Table VI. The parameter set 2 can be used for prediction of images of both sizes, 1024 × 768 and 1920 × 1088.

We analyze performances of the proposed DoC-DoGGRNN model using parameter set 2 by individual distortion types

applied to texture images or/and depth maps. We've used six subsets of the MCL-3D dataset, each with 108 images, synthesized from the content distorted by one of six distortion types: JPEG, JPEG-2000 compression, Gaussian blur, additive white noise, down-sampling blur and transmission error applied at four distortion levels. The proposed model achieves the best performance in evaluation of the images synthesized from the blurred content and compressed content, as shown in Table VII.

To further study model's performances in the evaluation of images synthesized from compressed content, we've used IST dataset [69] which contains images extracted from video sequences synthesized from texture/depth view pairs encoded with 3DV-ATM v10.0 [70], which is the 3D-AVC [71] reference software for MVD coding. IST dataset contains 180 images from 10 MVD sequences synthesized using both left

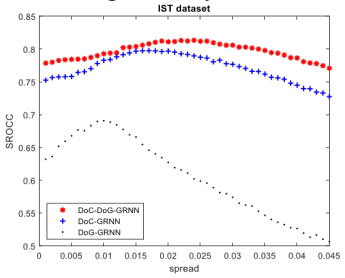


Fig. 8. SROCC of the proposed DoG-DoC-GRNN model, DoC-GRNN, DoG-GRNN models for IST image dataset.

TABLE VIII

PERFORMANCES OF THE PROPOSED DoC-DoG-GRNN MODEL USING PARAMETER SET 3 FOR IST DATASET AND PARTS OF THE MCL-3D DATASET

IST (180 images)			
	PLCC	SROCC	RMSE
Case 1	0.8278	0.8134	0.5561
Case 1A	0.8439	0.8135	0.5100
Case 2	0.8491	0.8367	0.5156
Case 2A	0.8522	0.8367	0.5084
Case 2B	0.8845	0.8656	0.4562
MCL-3D			
	PLCC	SROCC	RMSE
JPEG 2000 (108 images)	0.9600	0.9090	0.7232
JPEG (108 images)	0.9412	0.8519	0.7782

and right views compressed with different value of quantization parameter. The images have been synthesized by two synthesis algorithms, the VSRS-1D-Fast [72] and the VSIM [73]. Although the synthesized images contain both rendering and compression artifacts, the focus is on the distortion introduced by compressed texture and depth maps. To achieve computationally more efficient prediction of images synthesized from compressed content, we've selected parameter set 3: 4 resolution levels and 3 scales for DoC-based decomposition, $d_{i,1} - d_{i,3}$, i = 1,...,4, 4 resolution levels with single scale for DoG-based decomposition, $e_{i,1}$, i = 1,...,4, and spread parameter 0.022. Sparsity is calculated from 12 DoC bands, low-pass copy $s_{5,0}$ and 4 DoG bands. The GRNN input vector of the DoC-DoG-GRNN model is created by concatenation of 13 features calculated from DoC-based decomposition and 4 features calculated from DoG bands. The number of operation for calculation of 17 features of DoC-DoG-GRNN model using parameter set 3 is 10 times lower than for calculation of 46 features when parameter set 1 is used in unoptimized case. The SROCC of the composed DoC-DoG-GRNN model, DoC-GRNN and DoG-GRNN models with fluctuation of the GRNN spread parameter using DMOS as subjective scores for IST dataset, is shown on Fig. 8.

The performances of the DoC-DoG-GRNN model calculated using parameter set 3 evaluated on IST dataset and on two subsets of MCL-3D dataset, each with 108 stereo images synthesized from the content compressed by JPEG and JPEG 2000, calculated as case1, are shown in the Table VIII.

TABLE IX

PERFORMANCES OF THE PROPOSED MODEL AND OTHER METRICS FOR IST DATASET

IST dataset (180 images)				
	metric	PLCC	SROCC	RMSE
proposed model	**DoC-DoG-GRNN**	**0.8845**	**0.8656**	**0.4562**
blind IQA models for DIBR images	OMIQA [47]	0.2731	0.2899	0.9349
	APT [48]	0.1575	0.1446	0.9697
	NIQSV+ [50]	0.2348	0.2228	0.9447
	NIQSV [49]	0.3778	0.3517	0.8998
reference-based DIBR IQA models	MP-PSNR [18]	0.7322	0.7178	0.6619
	MW-PSNR [26]	0.7285	0.7103	0.6658
reference-based training-free IQA models for natural images	PSNR	0.6546	0.5843	0.7347
	SSIM	0.7030	0.6952	0.6911
	MS-SSIM [51]	0.8267	0.8115	0.5468

The model achieves high performances using parameter set 3 in the evaluation of images synthesized from compressed content with higher computational efficiency. Performances of the proposed model and other tested metrics evaluated on IST dataset are presented in Table IX. The performances of the proposed model are better than other tested metrics.

V. CONCLUSION

Considering the fact that image structures are crucial for visual quality perception and that the primary visual cortex and other parts of the brain reduce redundancies of input visual signals in order to discover the intrinsic image structure, thus creating sparse image representation, we propose the

referenceless model to evaluate the perceptual quality of DIBR-synthesized views. We define an oriented morphological Difference of Closings (DoC) operator and use it to nonlinearly remove redundancies and extract the higher-order structure (fine-grained image details-texture) of high curvature, which correspond to distortion of shapes. Using the DoC operator, perceptually important details in image local structure and contrast to which the HVS is highly sensitive are extracted at multiple scales and resolutions. We introduce a new feature based on sparsity of DoC band. To linearly remove redundancies and to extract the low-order structure (edges), the non-oriented Difference of Gaussians (DoG) operator is employed. Both operators, DoC and DoG, are applied at multiple scales and resolutions to mimic a hierarchical, multi-resolution and multi-scale character of HVS. Such two types of structural features play a complementary role in the visual quality assessment and they are sensitive to complex distortions of DIBR-synthesized views. Measures of sparsity of DoC and DoG bands are calculated as scalar features. Extracted features are mapped to the final score in a perceptually meaningful way by a trained general regression neural network. Performances of the proposed DoC-DoG-GRNN model that are calculated on four datasets of DIBR-synthesized views, IRCCyN/IVC image/video dataset, MCL-3D dataset and IST dataset demonstrate a high compatibility with perceptual quality scores, better than the tested state-ofthe-art models. The code for feature extraction is available at https://sites.google.com/site/draganasandicstankovic/code/doc-dog.

REFERENCES

[1] E. Bosc *et al.*, "Towards a new quality metric for 3-D synthesized view assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1332–1343, Nov. 2011.

[2] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jul. 1996.

[3] J. J. Koenderink, "The structure of images," *Biological*, vol. 50, no. 5, pp. 363–370, 1984.

[4] A. Dobbins, S. W. Zucker, and M. S. Cynader, "Endstopped neurons in the visual cortex as a substrate for calculating curvature," *Nature*, vol. 329, no. 6138, pp. 438–441, Oct. 1987.

[5] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.

[6] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Current Opin. Neurobiol.*, vol. 14, no. 4, pp. 481–487, Aug. 2004.

[7] A. Witkin, "Scale-space filtering: A new approach to multi-scale description," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 1984, pp. 150–153.

[8] K.-R. Park and C.-N. Lee, "Scale-space using mathematical morphology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 11, pp. 1121–1126, Nov. 1996.

[9] M.-H. Chen and P.-F. Yan, "A multiscanning approach based on morphological filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 694–700, Jul. 1989.

[10] P. Maragos, "Pattern spectrum and multiscale shape representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 701–716, Jul. 1989.

[11] P. Maragos and R. W. Schafer, "Morphological systems for multidimensional signal processing," *Proc. IEEE*, vol. 78, no. 4, pp. 690–710, Apr. 1990.

[12] F. Leymarie and M. Levine, "Shape features using curvature morphology," *Proc. SPIE*, vol. 1199, Nov. 1989, Art. no. 970050.

[13] D. Sandic´-Stankovi´c, D. M. Bokan, and D. D. Kukolj, "Blind DIBRsynthesized image quality assessment using multi-scale DoG and GRNN," in *Proc. 14th Symp. Neural Netw. Appl. (NEUREL)*, Nov. 2018, pp. 1–5.

[14] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, pp. 1457–1469, Dec. 2004.

[15] D. F. Specht, "A general regression neural network," *IEEE Trans. Neural Netw.*, vol. 2, no. 6, pp. 568–576, Nov. 1991.

[16] R. Song, H. Ko, and C. C. J. Kuo, "MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source," *J. Inf. Sci. Eng.*, vol. 31, no. 5, pp. 1593–1611, 2015.

[17] E. Bosc, P. Le Callet, L. Morin, and M. Pressigout, "Visual quality assessment of synthesized views in the context of 3D-TV," in *3D-TV System with Depth-Image-Based Rendering*, C. Zhu, Y. Zhao, L. Yu, and M. Tanimoto, Eds. New York, NY, USA: Springer, 2013.

[18] D. Sandic´-Stankovi´c, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological pyramids," in *Proc. 3DTV-Conf.: True Vis. - Capture, Transmiss. Display 3D Video (3DTVCON)*, Jul. 2015, pp. 1–4.

[19] D. Sandic´-Stankovi´c, D. Kukolj, and P. LeCallet, "Multi-scale synthesized view assessment based on morphological pyramids," *J. Elect. Eng.*, vol. 67, no. 1, pp. 1–9, 2016.

[20] J. Yan, Y. Fang, R. Du, Y. Zeng, and Y. Zuo, "No reference quality assessment for 3D synthesized views by local structure variation and global naturalness change," *IEEE Trans. Image Process.*, vol. 29, pp. 7443–7453, 2020.

[21] Y. Zhou, L. Yang, L. Li, K. Gu, and L. Tang, "Reduced-reference quality assessment of DIBR-synthesized images based on multi-scale edge intensity similarity," *Multimedia Tools Appl.*, vol. 77, pp. 21033–21052, Dec. 2017.

[22] L. Li, X. Chen, J. Wu, S. Wang, and G. Shi, "No-reference quality index of depth images based on statistics of edge profiles for view synthesis," *Inf. Sci.*, vol. 516, pp. 205–219, Apr. 2020.

[23] Y. Zhou, L. Li, S. Ling, and P. L. Callet, "Quality assessment for view synthesis using low-level and mid-level structural representation," *Signal Process.: Image Commun.*, vol. 74, pp. 309–321, May 2019.

[24] K. Gu, J.-F. Qiao, P. Le Callet, Z. Xia, and W. Lin, "Using multiscale analysis for blind quality assessment of DIBR-synthesized images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 745–749.

[25] K. Gu, J. Qiao, S. Lee, H. Liu, W. Lin, and P. Le Callet, "Multiscale natural scene statistical analysis for no-reference quality evaluation of DIBR-synthesized views," *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 127–139, Mar. 2019.

[26] D. Sandic´-Stankovi´c, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological wavelets," in *Proc. 7th Int. Workshop Quality Multimedia Exper. (QoMEX)*, Pilos, Greece, May 2015, pp. 1–6.

[27] D. Sandic´-Stankovi´c, D. Kukolj, and P. Le Callet, "DIBR-synthesized image quality assessment based on morphological multi-scale approach," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, p. 4, 2016.

[28] D. Sandic´-Stankovi´c, D. D. Kukolj, and P. Le Callet, "Fast blind quality assessment of DIBR-synthesized video based on high-high wavelet subband," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5524–5536, Nov. 2019.

[29] E. Bosc, F. Battisti, M. Carli, and P. L. Callet, "A wavelet-based image quality metric for the assessment of 3D synthesized views," *Proc. SPIE*, vol. 8648, Mar. 2013, Art. no. 86481Z.

[30] G. Wang, Z. Wang, K. Gu, L. Li, Z. Xia, and L. Wu, "Blind quality metric of DIBR-synthesized images in the discrete wavelet transform domain," *IEEE Trans. Image Process.*, vol. 29, pp. 1802–1814, 2019.

[31] Y. Zhou, L. Li, S. Wang, J. Wu, Y. Fang, and X. Gao, "No-reference quality assessment for view synthesis using DoG-based edge statistics and texture naturalness," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4566–4579, Sep. 2019.

[32] X. Wang, F. Shao, Q. Jiang, X. Meng, and Y.-S. Ho, "Measuring coarseto-fine texture and geometric distortions for quality assessment of

DIBRsynthesized images," *IEEE Trans. Multimedia*, vol. 23, pp. 1173–1186, 2021.

[33] S. Ling, J. Li, P. L. Callet, and J. Wang, "Perceptual representations of structural information in images: Application to quality assessment of synthesized view in FTV scenario," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1735–1739.

[34] S. Ling and P. Le Callet, "Image quality assessment for free viewpoint video based on mid-level contours feature," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 79–84.

[35] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, "Quality assessment of DIBR-synthesized images by measuring local geometric distortions and global sharpness," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 914–926, Apr. 2018.

[36] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 4, pp. 532–550, Jul. 1987.

[37] L. Vincent, "Fast grayscale granulometry algorithms," in *Proc. Int. Symp. Math. Morphol.*, Sep. 1994, pp. 265–272.

[38] R. A. Peters, "A new algorithm for image noise reduction using mathematical morphology," *IEEE Trans. Image Process.*, vol. 4, no. 5, pp. 554–568, May 1995.

[39] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, Feb. 1980.

[40] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009.

[41] D. Bokan, G. Velikic, D. Kukolj, and D. Sandic´-Stankovi´c, "Blind DIBR-synthesized image quality assessment based on sparsity features in morphological multiscale domain," in *Proc. 9th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2017, pp. 1–3.

[42] S. Ling and P. Le Callet, "How to learn the effect of non-uniform distortion on perceived visual quality? Case study using convolutional sparse coding for quality assessment of synthesized views," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018.

[43] *IRCCyN/IVC DIBR Image Quality Dataset*. Accessed: Dec. 11, 2013. [Online]. Available: ftp://ftp.ivc.polytech.univ-nantes.fr/IRCCyN_IVC_DIBR_Images

[44] *MCL-3D Stereoscopic Image Quality Dataset*. Accessed: Dec. 18, 2015. [Online]. Available: http://mcl.usc.edu/mcl-3d-database

[45] *IRCCyN/IVC DIBR Video Quality Dataset*. Accessed: Dec. 17, 2013. [Online]. Available: ftp://ftp.ivc.polytech.univnantes.fr/IRCCyN_IVC_DIBR_Videos

[46] *Test Plan for Evaluation of Video Quality Models for Use With High Definition TV Content*, VQEG HDTV Group, Boulder, CO, USA, 2009.

[47] V. Jakhetiya, K. Gu, T. Singhal, S. Guntuku, and W. Lin, "A highly efficient blind image quality assessment metric of 3-D synthesized images using outlier detection," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4120–4128, Jul. 2019.

[48] K. Gu, V. Jakhetiya, J. Qiao, X. Li, W. Lin, and D. Thalmann, "Modelbased referenceless quality metric of 3D synthesized images using local image description," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 394–405, Jan. 2018.

[49] S. Tian, L. Zhang, L. Morin, and O. Deforges, "NIQSV: A no reference image quality assessment metric for 3D synthesized views," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 1248–1252.

[50] S. Tian, L. Zhang, L. Morin, and O. Deforges, "NIQSV+: A noreference synthesized view quality assessment metric," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1652–1664, Apr. 2018.

[51] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, pp. 1398–1402.

[52] 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.

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

[54] W. Xue, L. Zhang, X. 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.

[55] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro- and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.

[56] 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. 4270–4281, Aug. 2014.

[57] A. Ahar, A. Barri, and P. Schelkens, "From sparse coding significance to perceptual quality: A new approach for image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 879–893, Feb. 2018.

[58] Y. Huang, X. Meng, and L. Li, "No-reference quality prediction for DIBR-synthesized images using statistics of fused color-depth images," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Aug. 2020, pp. 135–138.

[59] L. Wang, Y. Zhao, X. Ma, S. Qi, W. Yan, and H. Chen, "Quality assessment for DIBR-synthesized images with local and global distortions," *IEEE Access*, vol. 8, pp. 27938–27948, 2020.

[60] V. Jakhetiya, K. Gu, S. Jaiswal, T. Singhal, and Z. Xia, "Kernelridge regression-based quality measure and enhancement of threedimensional-synthesized images," *IEEE Trans. Ind. Electron.*, vol. 68, no. 1, pp. 423–433, Jan. 2021.

[61] S. Ling and P. Le Callet, "Image quality assessment for DIBR synthesized views using elastic metric," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 1157–1163.

[62] S. Ling, P. Le Callet, and G. Cheung, "Quality assessment for synthesized view based on variable-length context tree," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6. [63] S. Ling, P. Le Callet, and Z. Yu, "The role of structure and textural information in image utility and quality assessment tasks," *J. Perceptual Imag.*, vol. 2018, no. 1, p. 10501, Nov. 2018.

[64] S. Tian, L. Zhang, L. Morin, and O. Deforges, "A full-reference image quality assessment metric for 3D synthesized views," in *Proc. Int. Symp. Electron. Imag., Image Qual. Syst. Perform. (IS&T)*, Jan. 2018, p. 366.

[65] S. Tian *et al.*, "Quality assessment of DIBR-synthesized views: An overview," *Neurocomputing*, vol. 423, pp. 158–178, Jan. 2021.

[66] S. Ling, J. Li, Z. Che, X. Min, G. Zhai, and P. Le Callet, "Quality assessment of free-viewpoint videos by quantifying the elastic changes of multi-scale motion trajectories," *IEEE Trans. Image Process.*, vol. 30, pp. 517–531, 2021.

[67] S. Ling, J. Li, Z. Che, W. Zhou, J. Wang, and P. Le Callet, "Re-visiting discriminator for blind free-viewpoint image quality assessment," *IEEE Trans. Multimedia*, vol. 23, pp. 4245–4258, 2021.

[68] L. Krasula, K. Fliegel, P. Le Callet, and M. Klima, "On the accuracy of objective image and video quality models: New methodology for performance evaluation," in *Proc. 8th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Jun. 2016, pp. 1–6.

[69] F. Rodrigues, J. Ascenso, A. Rodrigues, and M. P. Queluz, "Blind quality assessment of 3-D synthesized views based on hybrid feature classes," *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1737–1749, Jul. 2019.

[70] *Reference Software for 3D-AVC: 3DV-ATM V10.0*. Accessed: Nov. 2013. [Online]. Available: http://mpeg3dv.nokiaresearch.com/ svn/mpeg3dv/tags/

[71] M. M. Hannuksela, Y. Chen, T. Suzuki, J. Ohm, and G. Sullivan, *3DAVC Draft Text 7, Joint Collaborative Team 3D Video Coding Extensions (JCT-3V)*, document JCT3V-E1002V2, Jan. 2013.

[72] *VSRS-1D-Fast*. Accessed: Nov. 2017. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware

[73] M. S. Farid, M. Lucenteforte, and M. Grangetto, "Depth image based rendering with inverse mapping," in *Proc. IEEE 15th Int. Workshop Multimedia Signal Process. (MMSP)*, Pula, Italy, Sep. 2013, pp. 135–140.

[74] M. Tanimoto, T. Fujii, and K. Suzuki, *View Synthesis Algorithm in View Synthesis Reference Software 3.5 (VSRS3.5)*, document M16090, ISO/IEC JTC1/SC29/WG11 (MPEG), May 2009.

[75] Y. J. Jung, H. G. Kim, and Y. M. Ro, "Critical binocular asymmetry measure for the perceptual quality assessment of synthesized stereo 3D images in view synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1201–1214, Jul. 2016.

[76] *Reference Software for Depth Estimation and View Synthesis*, document M15377, document ISO/IEC JTC1/SC29/WG11, Archamps, France, Apr. 2008.

[77] A. Criminisi, P. Pérez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.

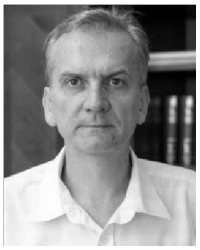
[78] I. Ahn and C. Kim, “A novel depth-based virtual view synthesis method for free viewpoint video,” *IEEE Trans. Broadcast.*, vol. 59, no. 4, pp. 614–626, Dec. 2013.

[79] S. S. Yoon, H. Sohn, Y. J. Jung, and Y. M. Ro, “Inter-view consistent hole filling in view extrapolation for multi-view image generation,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2883–2887.

[80] D. Bokan, D. Kukolj, and D. Sandic´-Stankovi´c, “A no-reference synthesized view quality assessment using statistical features in morphological multiscale domain,” in *Proc. 25th Telecommun. Forum (TELFOR)*, Nov. 2017, pp. 1–4.

[81] C. Li, A. C. Bovik, and X. Wu, “Blind image quality assessment using a general regression neural network,” *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 793–799, May 2011.

Dragana D. Sandic´-Stankovi´c received the M.Sc. degree in digital data transmission from the Faculty of Electrical Engineering, University of Belgrade, and the Ph.D. degree in electrical and computer engineering from the Faculty of Technical Sciences, University of Novi Sad, Serbia. She is currently a Researcher with the Institute IRITEL, Belgrade. She has authored more than 50 scientific and technical publications. Her research interests include the applications of multiresolution and multiscale decompositions using morphological filters in image and video processing and image and video quality assessment.



Dragan D. Kukolj (Senior Member, IEEE) received the Diploma degree in control engineering, the M.Sc. degree in computer engineering, and the Ph.D. degree in control engineering from the University of Novi Sad, Serbia, in 1982, 1988, and 1993, respectively. He is currently a Professor of computer-based systems with the Department of Computer Engineering, Faculty of Technical Sciences, University of Novi Sad. He has published more than 200 papers in referred journals and conference proceedings. His main research interests include machine learning, digital signal processing, and video processing.

Patrick Le Callet (Fellow, IEEE) is currently a Full Professor with Polytech Nantes, Nantes Université. 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 quality-of-experience assessment and visual attention modeling and applications. He is the coauthor of more than 300 publications and communications and a co-inventor of 16 international patents on these topics. He is a co-recipient of an Emmy Award in 2020 for his work on the development of perceptual metrics for video encoding optimization. He serves or has served as an Associate Editor or a Guest Editor for several journals such as *IEEE Signal Processing Magazine*, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and IEEE Transactions on Circuits and Systems for Video Technology.