The Visual Computer (2023) 39:6743–6753
https://doi.org/10.1007/s00371-022-02760-3

ORIGINAL ARTICLE

CoDIQE3D: A completely blind, no-reference stereoscopic image quality estimator using joint color and depth statistics

Ajay Kumar Reddy Poreddy[1] · Peter A. Kara[2] · Roopak R. Tamboli[3] · Aniko Simon[4] · Balasubramanyam Appina[5]

Accepted: 15 December 2022 / Published online: 4 January 2023

Abstract
In this paper, we present an unsupervised, completely blind, no-reference (NR) stereoscopic (S3D) image quality prediction model to assess the perceptual quality of natural S3D images. We study the joint dependencies between color and depth features of S3D images and empirically model these dependencies by using a bivariate generalized Gaussian distribution (BGGD). We compute the parameters of BGGD, and we also obtain the determinant and the coherence values from the covariance matrix of the proposed BGGD model. We extract the features of BGGD model and covariance matrix from the reference S3D image, followed by multivariate Gaussian (MVG) distribution modeling on the predicted features of the reference. We estimate the joint color and depth quality of the S3D images by computing the likelihood of the image features with respect to the reference MVG model. We apply the popular 2D unsupervised NIQE model on individual stereo views to estimate the overall spatial quality of the S3D images. Finally, we pool the likelihood scores and the spatial NIQE scores to achieve the estimation for the overall perceived quality of the S3D images. The performance of the proposed model is evaluated on the MICT, LIVE Phase I and II S3D image datasets. The results indicate consistent and robust performance for all datasets. Our proposed estimator is completely blind, as it requires neither training on subjective scores nor reference S3D images.

Keywords 3D image quality · Objective quality assessment · S3D · IQA · NIQE · Color · Depth

✉ Balasubramanyam Appina
appina@iiti.ac.in

Ajay Kumar Reddy Poreddy
edm20d012@iiitdm.ac.in

Peter A. Kara
kara@hit.bme.hu

Roopak R. Tamboli
ee13p0008@iith.ac.in

Aniko Simon
aniko.simon@sigmatechnology.com

[1] Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram, Chennai, India

[2] Department of Networked Systems and Services, Budapest University of Technology and Economics, Budapest, Hungary

[3] Department of Electrical Engineering, Indian Institute of Technology Hyderabad, Kandi, India

[4] Sigma Technology, Budapest, Hungary

[5] Department of Electrical Engineering, Indian Institute of Technology Indore, Simrol, Madhya Pradesh, India

1 Introduction

Among emerging 3D technologies, such as virtual reality (VR), augmented reality (AR) and light-field (LF) visualization, stereoscopic 3D (S3D) is an abundant source of many potential future use cases. A high portion of these use cases ambitiously target our everyday lives in our home environments. Unlike, for instance, LF, S3D technology is already entering households, somewhat slowly yet steadily. Still, according to the recent survey of Statista [1], the number of S3D displays globally purchased on the consumer market has reached 116,000 in 2019, which is roughly 40 times as many as measured in 2009.

S3D perception involves two individual 2D views and the corresponding depth. The source S3D content undergoes several processing stages, which may significantly affect the overall quality of experience (QoE) of the viewer. The impor-

tance of QoE must not be underestimated, as the satisfaction of the viewer always determines the true value of a system or service at the end of the day. However, the subjective evaluation of visual quality is a considerably expensive and lengthy process. Even though it is unquestionably the de facto method of assessing perceived quality in research, its utilization is far from being cost-efficient for regular usage, and it may also be subject to cognitive bias and other forms of rating distortion.

Objective quality assessment (QA) aims to overcome these unfavorable, disadvantageous properties of subjective evaluation by providing ratings without the need for human observers. Technically, objective QA estimators rate the quality of the given content automatically, but such methods differ in what they rely on when trying to complete this task. Therefore, this can be a base of classification among objective QA estimators. The scientific literature defines full reference (FR), reduced reference (RR) and no-reference (NR) objective metrics. In case of FR, the reference content—the quality of which is the best—is fully available to the estimator, and thus, a direct comparison with the investigated content is possible. RR metrics do not have complete access to the reference, and NR metrics basically have nothing to compare to, nothing to rely on.

Evidently, the most straightforward thing to do is to create FR metrics since their performance may easily outmatch RR and NR metrics due to their immense advantage. However, in many situations, the reference may not be fully available, and in fact, in even more cases, it may not be available at all. Consequently, NR metrics may fulfill their functionality of quality estimation under any given set of circumstances.

Color is an important visual cue for reconstructing both low-level and high-level visual features. The photoreceptors in the retinal ganglion cells capture rich information about the chromatic data in the visual cortex of the human visual system (HVS). Many authors conducted psychovisual studies [2,3] to measure the relative contribution of luminance and chromatic cues. They concluded that chromatic information could be used more efficiently than luminance information in solving the stereo correspondence issues. Later, Su et al. [4] modeled the conditional dependencies between luminance, color, and range data and found that the aforementioned cues exhibit a high correlation and also a high likelihood of joint dependencies between them. Appina et al. [5] proposed a supervised NR S3D IQA model based on the joint statistical dependencies between luminance and depth cues by using the bivariate generalized Gaussian distribution (BGGD) model. They weighted the model parameters of BGGD by using subband strengths of luminance and disparity cues. Finally, support vector regression (SVR) is employed to estimate the perceptual quality of S3D natural images. Motivated by the aforementioned studies, in this paper, we propose an unsupervised and completely blind (i.e., unaware of distortion and opinion) NR image quality assessment (IQA) model to evaluate the quality of S3D images. We parametrically model the joint statistical dependencies between color and depth cues by using a BGGD model and the spatial variations of the luminance cue by using the NIQE [6] model. The key technical contributions of the proposed model are the following:

1. We model the joint statistics of color and depth components by using a BGGD distribution and estimate the BGGD model features (α, β), coherence score (Ψ) and determinant measure (Δ) from the sample covariance matrix of the BGGD.
2. We compute the aforementioned features from the pristine stereoscopic image set and perform the multivariate Gaussian (MVG) distribution modeling on the estimated pristine features.
3. We estimate the joint color and depth quality of the S3D images by computing the likelihood of the image features with respect to the pristine MVG model.
4. We pool the joint color and depth quality scores with the spatial NIQE scores to estimate the perceptual quality of S3D images.

Technically, the previous work [5] used BGGD to model the joint statistics between luminance and depth features of S3D scenes and it did not consider any chromatic feature to estimate the subjective quality. Moreover, it did not utilize NIQE to achieve the results. In this work, we involve chromatic features—along with the depth and luminance features—to obtain the estimated scores. We model the joint dependencies between color and depth statistics with the BGGD model, and luminance quality is computed by performing the NIQE model on each view of an S3D image. Our method was tested on datasets commonly used within the scientific community for such purposes, and these promising results are presented and analyzed.

The remainder of the paper is structured as follows. An in-depth review of the related scientific literature is provided in Sect. 2. The detailed mechanisms of our proposed quality estimator are introduced in Sect. 3. The obtained results are presented and discussed in Sect. 4. The paper is concluded in Sect. 5.

2 Related work

S3D IQA models are more complex compared to traditional 2D IQA [6–12] methods due to the presence of depth scene and the quality of individual stereo views. Early works [13–16] introduced numerous FR S3D IQA algorithms based on computing the 2D IQA model performances. However, they conclude that 2D IQA models are not accurate in estimating the S3D image quality. Khan et al. and Bensalma et al. [5,17–21] proposed several stereo-

scopic IQA models by computing the features of sparsity, statistical, structural and binocular energy. Jiang et al. [22] published an FR S3D IQA method based on measuring the degradations in hierarchical features of 3D images by using the convolutional neural network (CNN) architecture.

In recent years, multiple studies [23–26] proposed NR S3D IQA models based on blurriness and blockiness estimates, saliency, and disparity structural properties. Shao et al. [27] designed an S3D IQA method by estimating the models of the phase-tuned visual codebook and lookup table of S3D images. The publications of Jiang et al. [28] and Chen et al. [29] presented NR IQA models based on saliency and scene statistical analysis on cyclopean image features. Zhou et al. [30] estimated binocular patterns and weights to create an NR S3D IQA algorithm.

Su et al. [31] performed univariate and bivariate analysis of scene statistics on marginal and adjacent subbands of the luminance component of a cyclopean image to estimate S3D image quality. Shi et al. [32] and Zhang et al. [33] proposed NR S3D IQA models based on CNN architectures. They computed features from individual stereo images and the corresponding difference image to predict the quality. Liu et al. [34] quantified the monocular and binocular properties of stereo scenes to achieve the goal of S3D image quality estimation.

Yildiz et al. [35] proposed a 3D FR IQA model to evaluate the quality of 3D static meshes. The model extracts the first four moments of the normalized histogram as descriptive features between the pristine and the test stimuli—degraded via the addition of noise and by smoothing the mesh with different strengths at four locations. Finally, a simple distance function is used between the pristine and the test features to evaluate the quality of the 3D static mesh.

In our earlier works, coordinated by the combined scientific efforts of Khan et al. [18] and Appina et al. [5], we modeled the marginal and joint luminance and depth subband coefficients with GGD and estimated the model features from GGD fits. Furthermore, we performed distance measurement and used the support vector regression (SVR) approach on computed features to estimate the quality of S3D images.

None of the aforementioned IQA algorithms and models studied or analyzed specifically the joint dependencies between color and depth components to estimate the S3D image quality. Also, these algorithms are not completely blind because they apply learning models on the estimated features and the corresponding subjective scores of the reference and the distorted S3D images. We propose a completely blind unsupervised NR S3D IQA model based on estimating the strength of joint dependencies between color and depth features of a stereoscopic image combined with NIQE quality measure. Again, **Co**lor and **D**epth statistics are used in our **I**mage **Q**uality **E**stimator for stereoscopic **3D** content, hence the name CoDIQE3D.

3 Proposed quality estimator

Various psychovisual experiments [36–39] were performed on the macaque visual cortex to analyze the selectivity and processing of color and depth components. They concluded that cortical neurons in the V2 and V3 areas are tuned for both color and depth components. Inspired by these experiments, Chen et al. and Su et al. [4,40–43] studied marginal and joint statistics of color and depth features of an S3D scene. They modeled the variation in Gabor responses of color and depth features and found that both the components are interdependent. We are motivated by the aforementioned psychovisual studies and experiments to model the joint dependencies between color and depth features of S3D images.

Though our algorithm is fundamentally inspired by our previous work [5], we strongly emphasize that none of the natural scene statistical analysis methods has modeled the joint dependencies between color and depth features of S3D images. Furthermore, we also highlight this model's utility by proposing a completely blind unsupervised NR IQA algorithm. We model the joint statistical dependencies between color and depth subband coefficients by using a BGGD, and feature parameters are computed from BGGD and the corresponding covariance matrix ($\mathbf{M}$). We show that the computed BGGD parameters and the coherence and determinant values of $\mathbf{M}$ clearly discriminate the distortions and utilize these features in estimating the quality of S3D images.

Consider a random vector $\mathbf{x} \in \mathbb{R}^N$ and the corresponding multivariate GGD (MGGD) [44] is given by

$$p(\mathbf{x}|\mathbf{M}, \alpha, \beta) = \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} g_{\alpha,\beta}(\mathbf{x}^T \mathbf{M}^{-1} \mathbf{x}), \tag{1}$$

$$g_{\alpha,\beta}(y) = \frac{\beta \Gamma\left(\frac{N}{2}\right)}{\left(2^{\frac{1}{\beta}} \Pi \alpha\right)^{\frac{N}{2}} \Gamma(\frac{N}{2\beta})} e^{-\frac{1}{2}(\frac{y}{\alpha})^{\beta}}, \tag{2}$$

where α and β are fitting parameters, and $\mathbf{M}$ is a covariance matrix with $N \times N$ dimension. In Eq. 2, if $\beta = \frac{1}{2}$, then the distribution is Laplacian, whereas for $\beta = 1$, the distribution is Gaussian. We performed maximum likelihood estimation to compute the model parameters (α, β, $\mathbf{M}$) of BGGD. We have color and depth as two features ($N = 2$), and hence, the above distribution turns into a BGGD. We utilize the BGGD model parameters (α, β), and determinant (Δ) and coherence (Ψ) values of $\mathbf{M}$ to estimate the quality. The coherence and determinant scores are defined as

$$\Delta = det(\mathbf{M}), \tag{3}$$

$$\Psi = \left[\frac{(\lambda_{\max} - \lambda_{\min})^2}{(\lambda_{\max} + \lambda_{\min})^2}\right], \tag{4}$$

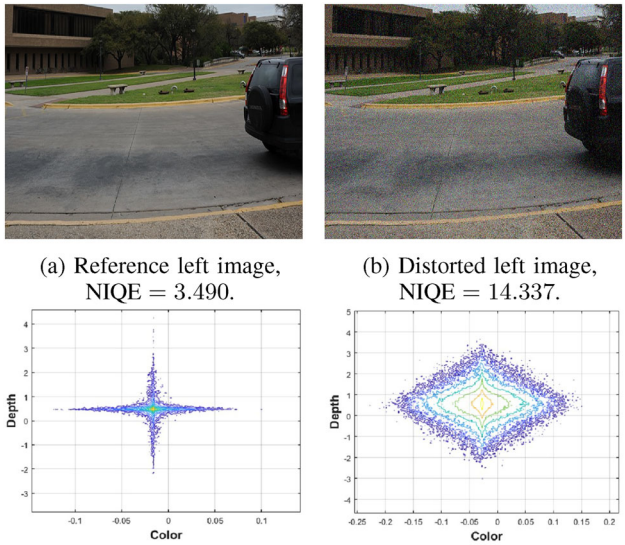
where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of $\mathbf{M}$. The eigenvalues measure the directionality of joint dependencies between color and depth features, and we are motivated by the work of Saad et al. [45] to compute the coherence score in the form (4). *det* represents the determinant of $\mathbf{M}$, and Δ is a dissimilarity measure between color and depth components. In our work, we perform multiple scale steerable pyramid decomposition [46] on color and depth maps. The bandpass filter response in a steerable subband decomposition mimics the properties of the human visual system (HVS). The prior work of Appina et al. [5] inspired us to perform three scale levels of steerable pyramid decomposition.

Figure 1a, b shows the left view of reference and the corresponding white Gaussian-impaired (WG-impaired) version of the *Car* S3D image [15], respectively. Figure 1c, d presents the joint contours between the color and depth subband coefficients of the reference and the distorted versions, respectively. The circular asymmetry nature in joint contour plots clearly demonstrates the strong reliances between color and depth features of S3D images, and we observe that such reliances vary with the type and level of the distortion. The noticeable variations observed in the joint contours indicate the change in BGGD parameters and $\mathbf{M}$. The change in $\mathbf{M}$ reflects the variation in the Δ and Ψ values. The joint contours in Fig. 1 are computed at the first scale of the steerable pyramid decomposition.

Figure 2a–d shows the variation of the α, β, Ψ and Δ scores at multiple subband decomposition of the reference and the corresponding WG noise distorted versions of the *Car* S3D image. These plots clearly demonstrate the efficacy of each estimated feature to discriminate the perceptual quality. We are motivated by these observations to use the aforementioned features of our proposed quality estimator. We perform negative logarithm computation on all features for better visualization in plots.

3.1 Algorithm overview

CoDIQE3D has four stages. In the first stage, we compute the color and depth features of an S3D image. The second stage involves BGGD modeling, followed by the 2D computation (CoDIQE2D). The third stage performs the spatial NIQE score estimation of the S3D image. In the fourth and final stage, we estimate the CoDIQE3D score by pooling the 2D and the spatial NIQE scores.



(a) Reference left image, NIQE = 3.490. (b) Distorted left image, NIQE = 14.337.

(c) $\alpha = 4 \times 10^{-7}$, β= 0.095, Ψ= 0.994, Δ= 0.422. (d) $\alpha = 2 \times 10^{-2}$, β= 0.457, Ψ= 0.961, Δ= 0.083.

Fig. 1 Illustration of joint contours between color and depth map of the reference and the distorted images. α, β, Ψ and Δ scores of the corresponding images are also visualized

3.2 Stage 1: feature extraction

3.2.1 Color feature set

In our metric, we perform the hue–saturation–value (HSV) color transformation on individual 2D RGB images. H and S measure the visual chrominance sensation and colorfulness of the image, respectively. Based on the study of Palm et al. [47], we compute the complex color representation from the H and S features.

$$\mathrm{Co}_L(m,n) = S_L(m,n) \times \exp^{(\mathrm{j}\times \mathrm{H_L(m,n)})}, \tag{5}$$

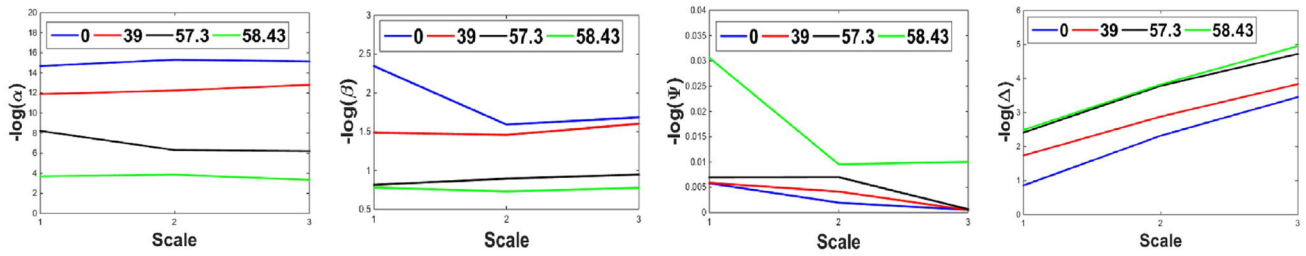
$$\mathrm{Co}_R(m,n) = S_R(m,n) \times exp^{(\mathrm{j}\times \mathrm{H_R(m,n)})}, \tag{6}$$

where S and H represent saturation and hue maps, j is an imaginary index, m and n are spatial coordinates, L and R are the individual left and right views of the S3D image, and Co represents the overall color map of the image.

Additionally, we consolidate individual color maps into a single color map of the S3D image. The entropy scores of Co_L and Co_R are estimated to pool the individual color maps.

$$C = \frac{\mathrm{ent}(\mathrm{Co}_L) \times \mathrm{Co}_L + \mathrm{ent}(\mathrm{Co}_R) \times \mathrm{Co}_R}{\mathrm{ent}(\mathrm{Co}_L) + \mathrm{ent}(\mathrm{Co}_R)}, \tag{7}$$

where C represents the overall color map of the S3D image, and *ent* is the entropy score of the color map. We consider this convex pooling strategy to mimic the properties of eye dominance in the HVS.



(a) α (b) β (c) Ψ (d) Δ

Fig. 2 Illustration of α, β, Ψ and Δ feature variation of the reference *Car* S3D image [15], and the corresponding versions distorted by white Gaussian noise. The corresponding difference mean opinion score (DMOS) is shown in the legend

The entropy *ent* denoted in the previous equation is calculated in the following way:

$$\text{ent} = -\sum_{a} p_a \times \log_2 p_a, \tag{8}$$

where a represents the number of bins and p denotes the normalized histogram counts of the image.

3.2.2 Depth feature set

We use a structural-similarity-based depth estimation algorithm [15] in our approach. The algorithm finds the best matching block between two stereo images by comparing the structural similarity between left and right views to estimate the corresponding depth map. The algorithm is robust and accurate with less time complexity. We perform multiple scale steerable pyramid subband decomposition on estimated color and depth maps.

3.3 Stage 2: CoDIQE2D computation

We perform BGGD fit on each subband coefficients of color and depth map of the S3D image and estimate the parameters of BGGD, determinant and coherence scores from the corresponding **M** of BGGD.

3.3.1 Reference MVG models

The IVY Lab dataset [48] has 120 reference S3D images, and we use these images to compute the reference MVG model parameters. We perform multiple scale steerable subband decomposition on the reference images and compute the BGGD parameters (α, β), Δ and Ψ values at each subband of the reference S3D image set. Mittal et al. [6] inspired us to model the individual features of the reference image set by using a multivariate Gaussian distribution and to estimate the mean vectors and covariance matrix of each feature α, β, Ψ and Δ, respectively.

3.3.2 Distorted feature set

We compute α, β, Ψ and Δ scores at multiple steerable subbands of the test image. We perform likelihood estimates of the computed test S3D image parameters (1×3 vector) for each feature coming from the reference MVG distribution.

$$T_\alpha = \mathcal{L}(\alpha_i^t; \alpha_\mu^r, \alpha_\Sigma^r), \tag{9}$$

$$T_\beta = \mathcal{L}(\beta_i^t; \beta_\mu^r, \beta_\Sigma^r), \tag{10}$$

$$T_\Psi = \mathcal{L}(\Psi_i^t; \Psi_\mu^r, \Psi_\Sigma^r), \tag{11}$$

$$T_\Delta = \mathcal{L}(\Delta_i^t; \Delta_\mu^r, \Delta_\Sigma^r), \tag{12}$$

where $\mathcal{L}$ is the likelihood function, r and t represent the reference and the test images, i is the subband level, T is the likelihood estimate, and μ and Σ are the mean and the covariance matrix of the reference MVG distribution of each reference feature α, β, Ψ and Δ.

The likelihood function used in the previous equations is the following:

$$\mathcal{L}(z; \mu, \sum) = \frac{1}{2\times(\pi)^{\frac{k}{2}} |\sum|^{\frac{1}{2}}} \times \exp\left(-\tfrac{1}{2}(z-\mu)' \sum^{-1} (z-\mu)\right), \tag{13}$$

where z is the distorted image feature set and k is the length of z. μ and $\sum$ are the mean and the covariance matrices of the reference MVG distribution.

We observe that T_α and T_β scores decrease and T_Ψ and T_Δ scores increase with distortion. These scores jointly affect the overall perceptual color and depth quality of the S3D image. Therefore, CoDIQE2D is defined as

$$\text{CoDIQE2D} = \exp\left(\frac{T_\alpha \times T_\beta}{T_\Psi \times T_\Delta}\right), \tag{14}$$

where CoDIQE2D is the overall color and depth quality of the S3D image. We apply exp on the product score to mimic the numerical range of spatial quality scores.

Table 1 shows the efficacies of the individual T_α, T_β, T_Ψ, T_Δ and the proposed CoDIQE2D pooling on the LIVE Phase II S3D image dataset. The results clearly demonstrate the

Table 1 The proposed CoDIQE2D and the corresponding individual feature performances on the LIVE Phase II dataset

	T_α	T_β	T_Ψ	T_Δ	CoDIQE2D
LCC	0.2923	0.2526	0.1951	0.1454	0.4571
SROCC	0.2410	0.2370	0.1827	0.1101	0.4095

significance of each feature in estimating the overall color and depth quality of the S3D image.

3.4 Stage 3: Spatial feature

We compute the popular unsupervised NR 2D IQA model NIQE [6] on individual stereo views. The overall spatial quality of an S3D image is calculated by computing the mean of the individual NIQE scores.

$$\text{Spa} = \frac{\text{NIQE}_L + \text{NIQE}_R}{2}, \tag{15}$$

where NIQE_L, NIQE_R represent the individual spatial NIQE scores, while *Spa* denotes the overall spatial quality score of the S3D image.

The NIQE calculation included in the previous equation is performed in the following manner:

$$\text{NIQE}_L = \text{niqe(L)}, \ \ \text{NIQE}_R = \text{niqe(R)}, \tag{16}$$

where *niqe* represents the natural image quality evaluator algorithm.

3.5 Stage 4: CoDIQE3D computation

The overall color and depth quality (CoDIQE2D) scores decrease, and spatial quality (Spa) scores increase with test stimuli strength. Both scores jointly affect the final perceptual quality estimate of the S3D image. Therefore, we compute the product between 1/CoDIQE2D and *Spa* scores to derive the CoDIQE3D score of the S3D image:

$$\text{CoDIQE3D} = \text{Spa} \times \frac{1}{\text{CoDIQE2D}} \tag{17}$$

4 Results and discussion

The performance of the proposed CoDIQE3D model is evaluated on the MICT [23] and LIVE Phase I and II datasets [15,49]. The MICT dataset consists of 13 pristine reference and 624 JPEG-distorted images. The LIVE Phase I and II datasets are composed of images distorted with blur, fast fading (FF), JP2K, JPEG and WG. The LIVE Phase I dataset consists of 20 reference and 365 distorted versions. The LIVE Phase II dataset has 8 reference and 360 distorted versions. From these 360 distorted versions, 240 are symmetrically and 120 are asymmetrically distorted. As a base of comparison for the human perception of quality, DMOS scores are used. To report the performance of the proposed model, LCC, SROCC and RMSE statistics are measured between the CoDIQE3D predictions and the actual DMOS scores. The performance numbers above are reported after performing a nonlinear logistic fit [51].

Table 2 shows the results of the evaluation of the proposed CoDIQE3D model on the LIVE Phase I and II datasets. It is clear that CoDIQE3D demonstrates consistent and robust performance over different test stimuli types. Figure 4 shows scatter plots of *Spa* and CoDIQE3D on the LIVE Phase II dataset. Furthermore, these scatter plots provide an additional evidence about the efficiency of the model.

Tables 3 and 4 show the performance comparison of the CoDIQE3D model with off-the-shelf 2D and 3D IQA models on LIVE Phase I, MICT and LIVE Phase II S3D IQA datasets. SSIM [7], MS-SSIM [8], NIQE [6] and IL-NIQE [10] 2D IQA models are applied on individual views, and the mean of both view scores is computed to represent the S3D image quality. Chen [15], Bensalma [20], Benoit [16], Striqe [18], Wang [50], Shao [27], Chen [29], StereoQue [5], S3DBlinque [31], BSIQE [34] and RM-CNN3 [32] are FR and supervised NR S3D IQA models. The proposed model shows robust performance across all distortions and consistent performance across all S3D IQA datasets. It also shows competitive performance against 2D and 3D FR IQA models and 3D supervised NR IQA models. The CoDIQE3D model delivers state-of-the-art performance when compared with unsupervised IQA models. To the best of our knowledge, CoDIQE3D is the only model to estimate the quality of S3D natural images in an unsupervised theme.

Furthermore, we investigate the efficacy of CoDIQE3D by replacing the spatial NIQE model with another popular 2D unsupervised NR IQA model, the IL-NIQE. Table 5 shows the performance of the model on the LIVE Phase I and II, and MICT datasets, by using 2D unsupervised NR IQA models. It is apparent that CoDIQE3D shows improved performance over 2D unsupervised NR IQA models across the datasets. This analysis provides additional supporting evidence of the proposed CoDIQE2D feature for adequate color and depth quality representation.

Blur and fast fading distortions highly affect the color and depth perception of a stereoscopic scene [15,40], and it reflects more variation in joint dependencies compared to other distortions. The proposed CoDIQE3D algorithm efficiently captures these statistical variations between color and depth features of an S3D scene and demonstrates better performance on blur and fast fading distortions compared to other types of visual distortions.

Additionally, while the 2D unsupervised IQA models NIQE and IL-NIQE perform rather well on symmetrically distorted images, they do fail to reproduce the same performance on asymmetrically distorted images (as shown in Table 3). In contrast, CoDIQE3D—pooling the quality scores of joint color and depth, and spatial quality—shows consistent and improved performance on both distortion types (i.e., symmetric and asymmetric). We attribute this robust execution of the quality prediction task to the choice of the distortion discriminable features.

In addition to the evaluation, we analyzed the performance of the proposed CoDIQE3D model on the synthetic SELMA dataset [52], which was developed for semantic segmentation. The dataset contains more than 30,000 images. These images are acquired by using 24 sensors in different times of the day, and multiple weather conditions are simulated by using the well-known CARLA [53] software. The images are provided in four formats: RGB, depth, semantic and LiDAR.

To demonstrate the efficacy of the CoDIQE3D model on these synthetic images, we modeled the joint dependencies between the RGB and depth planes of the SELMA dataset by using a BGGD model. Figure 3 represents the distribution of the α, β, Δ and Ψ scores of the 'Town06' image of the SELMA dataset at six different day times in a rainy setting, by using a 2D tSNE plot [54]. This plot is computed between the first two components of the tSNE matrix. The achieved output clearly indicates that the parameters of the CoDIQE3D model are clustered quite well with respect to different day times in the selected rainy day scenario.

However, going beyond these results and performing a comprehensive evaluation in the same fashion as on the other datasets is not feasible, as the authors did not conduct any subjective analysis, and therefore, the actual performance of the CoDIQE3D model on the SELMA dataset cannot be assessed.

Table 2 The proposed CoDIQE3D performance evaluation on the LIVE Phase I [49] and II [15] S3D image datasets

Distortion	Phase I dataset			Phase II dataset		
	LCC	SROCC	RMSE	LCC	SROCC	RMSE
JP2K	0.532	0.483	14.261	0.700	0.672	7.001
BLUR	0.932	0.881	5.220	0.935	0.874	5.033
FF	0.740	0.705	11.083	0.871	0.865	5.251
JPEG	0.474	0.448	15.755	0.697	0.637	6.020
WN	0.605	0.582	13.235	0.852	0.847	5.253
Overall	0.778	0.769	10.586	0.790	0.766	7.189

Table 3 Performance comparison of CoDIQE3D with popular 2D&3D IQA models on the LIVE Phase II dataset [15]

Model type	Algorithm	Symm		Asymm		Overall	
		LCC	SROCC	LCC	SROCC	LCC	SROCC
2D FR IQA	SSIM [7]	0.733	0.700	0.750	0.719	0.873	0.877
	MS-SSIM [8]	0.924	0.912	0.709	0.687	0.780	0.784
2D NR IQA unsupervised	NIQE [6]	0.771	0.737	0.701	0.671	0.732	0.710
	IL-NIQE [10]	0.739	0.697	0.630	0.603	0.727	0.700
3D FR IQA	Chen [15]	0.938	0.923	0.875	0.842	0.900	0.889
	Bensalma [20]	0.923	0.805	0.743	0.697	0.770	0.751
	Benoit [16]	0.734	0.696	0.770	0.747	0.762	0.744
	Striqe [18]	0.909	0.910	0.889	0.868	0.901	0.892
	Wang [50]	0.937	0.923	0.898	0.902	0.915	0.918
3D NR IQA supervised	Shao [27]	0.911	0.896	0.565	0.524	–	–
	Chen [29]	–	0.918	–	0.834	0.895	0.880
	StereoQue [5]	–	0.857	–	0.872	0.845	0.888
	S3DBlinque [31]	–	0.937	–	0.849	0.913	0.905
	BSIQE [34]	–	0.937	–	0.903	0.934	0.929
	RM-CNN3 [32]	0.970	0.928	0.953	0.943	0.961	0.948
	CoDIQE2D	0.521	0.404	0.377	0.318	0.457	0.409
	CoDIQE3D	0.829	0.792	0.756	0.718	0.790	0.766

Table 4 Performance comparison of CoDIQE3D with popular 2D&3D IQA models on the LIVE Phase I [49] and MICT datasets [23]

Model type	Algorithm	LIVE Phase I		MICT	
		LCC	SROCC	LCC	SROCC
2D FR IQA	SSIM [7]	0.872	0.876	0.862	0.846
	MS-SSIM [8]	0.926	0.926	0.935	0.935
2D NR IQA Unsupervised	NIQE [6]	0.752	0.735	0.778	0.761
	IL-NIQE [10]	0.740	0.705	0.725	0.695
3D FR IQA	Chen [15]	0.916	0.916	0.864	0.862
	Bensalma [20]	0.887	0.875	0.954	0.940
	Benoit [16]	0.902	0.728	0.901	0.902
	Striqe [18]	0.927	0.922	0.772	0.768
	Wang [50]	0.924	0.916	–	–
3D NR IQA supervised	Shao [27]	0.904	0.875	–	–
	Chen [29]	0.895	0.891	0.913	0.904
	StereoQue [5]	0.917	0.888	0.935	0.936
	S3DBlinque [31]	0.895	0.886	0.933	0.917
	BSIQE [34]	0.959	0.949	–	–
	RM-CNN3 [32]	0.960	0.936	–	–
	CoDIQE2D	0.495	0.453	0.419	0.386
	CoDIQE3D	0.778	0.769	0.786	0.776

Table 5 Performance evaluation of CoDIQE3D on different datasets with popular 2D unsupervised NR IQA algorithms

Dataset	NIQE		IL-NIQE	
	LCC	SROCC	LCC	SROCC
LIVE Phase I	0.778	0.769	0.746	0.717
LIVE Phase II	0.790	0.766	0.771	0.740
MICT	0.786	0.776	0.731	0.700

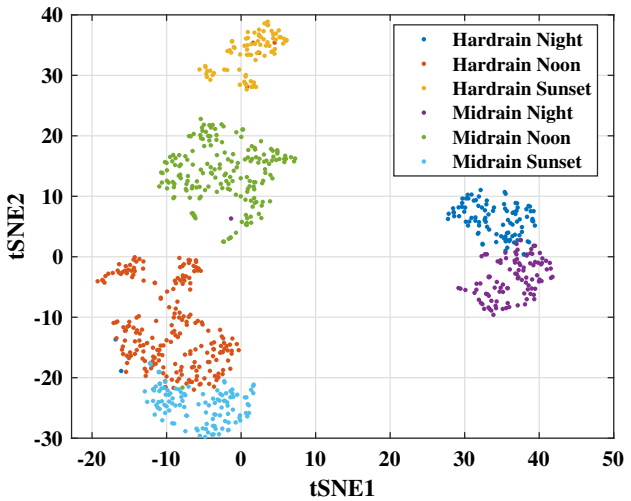


Fig. 3 Illustration of the tSNE plot between the parameters of the proposed CoDIQE3D model on six different day times of the rainy weather conditions of the 'Town06' image of the SELMA dataset

According to the best knowledge of the authors, CoDIQE3D is the first and only estimator to perform statistical analysis on the dependencies between color and depth subband coefficients of natural S3D images. Again, the algorithm follows a completely blind approach and does not require training on stereoscopic image features and the corresponding annotations. Such quality estimators are rather relevant to the multitude of contexts in which S3D quality plays a significant role. These evidently include the broadcasting of S3D contents, which may be affected by unfavorable transmission parameters, leading to suboptimal levels of perceived quality. With objective metrics, such as the one presented in this paper, the QoE of the users of S3D systems and services may be cost-efficiently modeled.

5 Conclusion

In this paper, we proposed a completely blind unsupervised NR IQA model based on estimating the joint dependency strength between color and depth features of S3D images. We computed the parameters (α, β) of BGGD and determinant (Δ) and coherence (Ψ) values from the covariance matrix and demonstrated the efficacy of these features in distortion discrimination. Furthermore, an unsupervised NR IQA NIQE model was utilized to compute the overall spatial quality of the images. Our results clearly indicate that the proposed CoDIQE3D model, which had no prior training or testing on distorted features with associated quality scores, demonstrates robust performance and competitive-

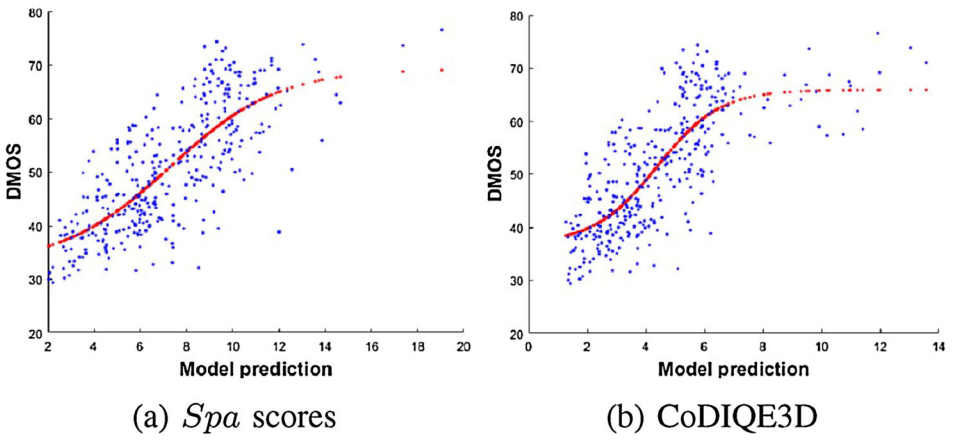


Fig. 4 Scatter plots of *Spa* and CoDIQE3D objective scores versus DMOS on the LIVE Phase II dataset

ness against popular FR and supervised NR 2D and 3D IQA models.

In the future, we plan to extend the proposed method to S3D video quality assessment. Although such extension could be performed in a straightforward manner by simply applying CoDIQE3D to each individual frame, involving interframe redundancies could further improve the accuracy of prediction. Another potential future continuation of this work would be the adaptation of the applied concepts of CoDIQE3D to emerging 3D visualization technologies, such as light field. By doing so, an evaluator designed for stereoscopic 3D utilization would benefit the quality assessment of autostereoscopic 3D technologies.

Acknowledgements The research reported in this paper was supported in part by the Department of Science and Technology - Science and Engineering Research Board, Government of India under Grant SRG/2020/000336. The work was also supported by project no. BME-NVA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme.

Data availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Statista (Julia Stoll): Number of digital 3D cinema screens worldwide from 2006 to 2019. https://www.statista.com/statistics/271863/number-of-3d-cinema-screens-worldwide/ (2020)
2. Jordan, J.R., III., Bovik, A.C.: Using chromatic information in edge-based stereo correspondence. CVGIP Image Underst. **54**(1), 98–118 (1991)
3. Jordan, J.R., III., Bovik, A.C.: Using chromatic information in dense stereo correspondence. Pattern Recogn. **25**(4), 367–383 (1992)
4. Su, C.-C., Cormack, L.K., Bovik, A.C.: Color and depth priors in natural images. IEEE Trans. Image Process. **22**(6), 2259–2274 (2013)
5. Appina, B., Khan, S., Channappayya, S.S.: No-reference stereoscopic image quality assessment using natural scene statistics. Signal Process.: Image Commun. **43**, 1–14 (2016)
6. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. IEEE Signal Process. Lett. **20**(3), 209–212 (2013)
7. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)
8. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: Conference on Signals, Systems and Computers, vol. 2, pp. 1398–1402, IEEE (2003)
9. Cai, R., Fang, M.: Blind image quality assessment by simulating the visual cortex. Vis. Comput. 1–18 (2022)
10. Zhang, L., Zhang, L., Bovik, A.C.: A feature-enriched completely blind image quality evaluator. IEEE Trans. Image Process. **24**(8), 2579–2591 (2015)
11. Ji, J., Xiang, K., Wang, X.: SCVS: blind image quality assessment based on spatial correlation and visual saliency. Vis. Comput. **1**, 23 (2022)
12. Joshi, P., Prakash, S., Rawat, S.: Continuous wavelet transform-based no-reference quality assessment of deblocked images. Vis. Comput. **34**(12), 1739–1748 (2018)
13. Campisi, P., Le Callet, P., Marini, E.: Stereoscopic images quality assessment. In: European Signal Processing Conference, pp. 2110–2114. IEEE (2007)
14. Gorley, P., Holliman, N.: Stereoscopic image quality metrics and compression. In: Electronic Imaging, pp. 45–56. International Society for Optics and Photonics (2008)
15. Chen, M.-J., Su, C.-C., Kwon, D.-K., Cormack, L.K., Bovik, A.C.: Full-reference quality assessment of stereopairs accounting for rivalry. Signal Process.: Image Commun. **28**(9), 1143–1155 (2013)
16. Benoit, A., Le Callet, P., Campisi, P., Cousseau, R.: Quality assessment of stereoscopic images. EURASIP J. Image Video Process. **2008**, 1–13 (2009)
17. Khan, M.S., Channappayya, S.S.: Sparsity based stereoscopic image quality assessment. In: Asilomar Conference on Signals, Systems and Computers, pp. 1858–1862. IEEE (2016)
18. Khan, Md.S., Appina, B., Channappayya, S.: Full-reference stereo image quality assessment using natural stereo scene statistics. IEEE Signal Process. Lett. **22**, 1985–1989 (2015)

19. Khan, S., Channappayya, S.S.: Estimating depth-salient edges and its application to stereoscopic image quality assessment. IEEE Trans. Image Process. **27**(12), 5892–5903 (2018)
20. Bensalma, R., Larabi, M.-C.: A perceptual metric for stereoscopic image quality assessment based on the binocular energy. Multidimens. Syst. Signal Process. **24**(2), 281–316 (2013)
21. Bensalma, R., Larabi, M.C.: Towards a perceptual quality metric for color stereo images. In International Conference on Image Processing, pp. 4037–4040. IEEE (2010)
22. Jiang, Q., Zhou, W., Chai, X., Yue, G., Shao, F., Chen, Z.: A full-reference stereoscopic image quality measurement via hierarchical deep feature degradation fusion. IEEE Trans. Instrum. Meas. **69**(12), 9784–9796 (2020)
23. Akhter, R., Sazzad, Z.P., Horita, Y., Baltes, J.: No-reference stereoscopic image quality assessment. In: IS&T/SPIE Electronic Imaging, pp. 271–282. International Society for Optics and Photonics (2010)
24. Sazzad, Z.P., Yamanaka, S., Kawayokeita, Y., Horita, Y.: Stereoscopic image quality prediction. In: International Workshop on Quality of Multimedia Experience, pp. 180–185, IEEE (2009)
25. Ryu, S., Sohn, K.: No-reference quality assessment for stereoscopic images based on binocular quality perception. IEEE Trans. Circuits Syst. Video Technol. **24**(4), 591–602 (2013)
26. Appina, B.: A 'Complete Blind' No-Reference Stereoscopic Image Quality Assessment Algorithm. In: International Conference on Signal Processing and Communications (SPCOM), pp. 1–5. IEEE (2020)
27. Shao, F., Lin, W., Wang, S., Jiang, G., Yu, M.: Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook. IEEE Trans. Broadcast. **61**, 154–165 (2015)
28. Jiang, Q., Duan, F., Shao, F.: 3D visual attention for stereoscopic image quality assessment. J. Softw. **9**(7), 1841–1847 (2014)
29. Chen, M.-J., Cormack, L.K., Bovik, A.C.: No-reference quality assessment of natural stereopairs. IEEE Trans. Image Process. **22**(9), 3379–3391 (2013)
30. Zhou, W., Yu, L., Zhou, Y., Qiu, W., Wu, M.-W., Luo, T.: Blind quality estimator for 3D images based on binocular combination and extreme learning machine. Pattern Recogn. **71**, 207–217 (2017)
31. Su, C.-C., Cormack, L.K., Bovik, A.C.: Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation. IEEE Trans. Image Process. **24**(5), 1685–1699 (2015)
32. Shi, Y., Guo, W., Niu, Y., Zhan, J.: No-reference stereoscopic image quality assessment using a multi-task CNN and registered distortion representation. Pattern Recogn. **100**, 1–12 (2020)
33. Zhang, W., Qu, C., Ma, L., Guan, J., Huang, R.: Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network. Pattern Recogn. **59**, 176–187 (2016)
34. Liu, Y., Yan, W., Zheng, Z., Huang, B., Yu, H.: Blind stereoscopic image quality assessment accounting for human monocular visual properties and binocular interactions. IEEE Access **8**, 33666–33678 (2020)
35. Yildiz, Z.C., Oztireli, A.C., Capin, T.: A machine learning framework for full-reference 3D shape quality assessment. Vis. Comput. **36**(1), 127–139 (2020)
36. Den Ouden, H., Van Ee, R., De Haan, E.: Colour helps to solve the binocular matching problem. J. Physiol. **567**(2), 665–671 (2005)
37. Daniel, Y., Roe, A.W., Gilbert, C.D.: A hierarchy of the functional organization for color, form and disparity in primate visual area v2. Vision. Res. **41**(10–11), 1333–1349 (2001)
38. Nasr, S., Polimeni, J.R., Tootell, R.B.: Interdigitated color-and disparity-selective columns within human visual cortical areas v2 and v3. J. Neurosci. **36**(6), 1841–1857 (2016)
39. Fine, I., MacLeod, D., Boynton, G.M.: Surface segmentation based on the luminance and color statistics of natural scenes. J. Vis. **2**(10), 66–66 (2002)
40. Chen, M., Bovik, A.C., Cormack, L.K.: Study on distortion conspicuity in stereoscopically viewed 3D images. In: IEEE 10th IVMSPWorkshop: Perception and Visual Signal Analysis, pp. 24–29 (2011)
41. Su, C.C., Bovik, A.C., Cormack, L. K.: Statistical model of color and disparity with application to Bayesian stereopsis. In: Southwest Symposium on Image Analysis and Interpretation, pp. 169–172. IEEE (2012)
42. Su, C.C., Bovik, A.C., Cormack, L.K.: Natural scene statistics of color and range. In: International Conference on Image Processing, pp. 257–260, IEEE (2011)
43. Su, C.-C., Cormack, L.K., Bovik, A.C.: Bivariate statistical modeling of color and range in natural scenes. In: IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, pp. 391–400 (2014)
44. Pascal, F., Bombrun, L., Tourneret, J.-Y., Berthoumieu, Y.: Parameter estimation for multivariate generalized Gaussian distributions. IEEE Trans. Signal Process. **61**(23), 5960–5971 (2013)
45. Saad, M.A., Bovik, A.C., Charrier, C.: Blind prediction of natural video quality. IEEE Trans. Image Process. **23**(3), 1352–1365 (2014)
46. Simoncelli, E.P., Freeman, W.T.: The steerable pyramid: a flexible architecture for multi-scale derivative computation. In: Proceedings., International Conference on Image Processing, vol. 3, pp. 444–447. IEEE (1995)
47. Palm, C., Keysers, D., Lehmann, T., Spitzer, K.: Gabor filtering of complex hue/saturation images for color texture classification. In: Proceeding of the JCIS, pp. 45–49, Citeseer (2000)
48. Jung, Y.J., Sohn, H., Lee, S.-I., Park, H.W., Ro, Y.M.: Predicting visual discomfort of stereoscopic images using human attention model. IEEE Trans. Circuits Syst. Video Technol. **23**(12), 2077–2082 (2013)
49. Moorthy, A.K., Su, C.-C., Mittal, A., Bovik, A.C.: Subjective evaluation of stereoscopic image quality. Signal Process. Image Commun. **28**(8), 870–883 (2013)
50. Wang, J., Rehman, A., Zeng, K., Wang, S.,Wang, Z.: Quality prediction of asymmetrically distorted stereoscopic 3D images. IEEE Trans. Image Process. 24(11), 3400–3414 (2015)
51. VQEG Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II. [online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-ii/frtv-phase-ii.aspx (2003)
52. Testolina, P., Barbato, F., Michieli, U., Giordani, M., Zanuttigh, P., Zorzi, M.: SELMA: SEmantic large-scale multimodal acquisitions in variable weather, daytime and viewpoints. arXiv preprint arXiv:2204.09788 (2022)
53. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: an open urban driving simulator. In: Conference on Robot Learning, pp. 1–16. PMLR (2017)
54. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(11), 2579–2605 (2008)

Ajay Kumar Reddy Poreddy is currently a Research Scholar in Electronics and Communication Engineering at Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram. He is doing research in 2D&3D image and video quality assessment using natural scene statistical (NSS) features.

Peter A. Kara received his MSc degree in Computer Engineering from the Department of Networked Systems and Services (HIT) at the Budapest University of Technology and Economics (BME) in 2013, and he was awarded the PhD title by Kingston University (KU) in 2020. He participated in the EU FP7 ICT CONCERTO and EU H2020 QoE-NET projects, worked as a research associate at KU, and he is currently a research fellow of the Multimedia Networks and Services Laboratory (MEDIANETS) at BME-HIT, an invited fellow of KU and vice-chair of IEEE P3333.1.4.

Roopak R. Tamboli recently received his PhD degree from the Dept. of Electrical Engineering, Indian Institute of Technology Hyderabad. He works in the areas of 3D reconstruction from multiple views and quality assessment of super-multiview (SMV) content. In particular, his work includes generation of multiview content, true-to-scale 3D reconstruction from 2D views, subjective and objective methods for assessing perceptual quality of SMV content visualized on projection-based light field (LF) displays.

Aniko Simon received her MSc degree in Computer Engineering from the Budapest University of Technology and Economics (BME) in 2015. During her studies at BME, she participated in the research and development efforts of the Mobile Innovation Centre (MIK). She is an expert in technical communication, and she is currently an information engineer at Sigma Technology.

Balasubramanyam Appina is an Assistant professor in the department of Electrical Engineering at Indian Institute of Technology Indore, India. He received M.Tech and PhD degrees in Electrical Engineering from Indian Institute of Technology Hyderabad in 2015 and 2019, respectively. His research interests include multimedia quality assessment, image and video processing and display technology. He was a recipient of the Young professional grant award from Electronic Imaging Society in 2017.