On the practical applications of objective quality metrics for stereoscopic 3D imaging

Sria Biswas[a], Balasubramanyam Appina[a], Roopak R. Tamboli[b],
Peter A. Kara[c], and Aniko Simon[d]

[a]Indian Institute of Information Technology, Design and Manufacturing Kancheepuram, Chennai, India
[b]Saarland University, Saarbrücken, Germany
[c]Budapest University of Technology and Economics, Budapest, Hungary
[d]Sigma Technology, Budapest, Hungary

ABSTRACT

Objective quality metrics provide cost-efficient methods for quality evaluation, as they are practically algorithms, models, that avoid the necessity of subjective assessment, which is a precise but resource-consuming approach. Their ultimate measure of prediction accuracy fundamentally relies on the correlation between the estimated levels of quality and the actual subjective scores of perceived quality, rated by human individuals. Such metrics have already been developed for every single emerging technology where quality, in general, is relevant. This applies to stereoscopic 3D imaging as well, which is utilized in both industry, healthcare, education and entertainment. In this paper, we introduce an exhaustive analysis regarding the practical applications of objective quality metrics for stereoscopic 3D imaging. Our contribution addresses each and every state-of-the-art objective metric in the scientific literature, separately for image and video quality. The study differentiates the metrics by input requirements and supervision, and examines performance via statistical measures. Machine learning algorithms are particularly emphasized within the paper, such as the Deep Edge and COlor Signal INtegrity Evaluator (DECOSINE) using Segmented Stacked Auto-Encoder (S-SAE), different Convolutional Neural Network (CNN) frameworks, and transfer-learning-based methods like the Xception model, AlexNet, ResNet-18, ImageNet, Caffe, GoogLeNet, and also our very own transfer-learning-based methods. The paper focuses on the actual practical applications of the predictive models, and highlights relevant criteria, along with general feasibility, suitability and usability. The analysis of the investigated use cases also addresses potential future research questions and specifies the appropriate directives for quality-focused, user-centric development.

Keywords: Image quality assessment, video quality assessment, stereoscopic 3D, machine learning, deep learning

1. INTRODUCTION

The recent advancements in stereoscopic 3D (S3D) image and video processing have led to its booming demands in the entertainment, education and healthcare industries. The rapid development of S3D technology resulted in the need for more effective, accurate and efficient image quality assessment (IQA) and video quality assessment (VQA) techniques. Earlier works in the field of IQA and VQA mainly dealt with 2D technologies which considered spatial and temporal information only. Stereoscopy takes into account the depth information as well, which is vital to the human visual system (HVS), and thus, leads to better Quality of Experience (QoE). The human eye

Further author information: (Send correspondence to Balasubramanyam Appina)
Sria Biswas: E-mail: ec21d0002@iiitdm.ac.in
Balasubramanyam Appina: E-mail: appina@iiitdm.ac.in
Roopak R. Tamboli: E-mail: tamboli@nt.uni-saarland.de
Peter A. Kara: E-mail: kara@hit.bme.hu
Aniko Simon: E-mail: aniko.simon@sigmatechnology.se

perceives the visual world in 3D binocular view since it receives visual stimuli through both eyes and then fuses them into one single perception, hence stereoscopic techniques consider the left and right 2D monocular views and overlap them into a virtual image placed in between these two views called the Cyclopean image.[1] This cyclopean image view is an important part in many stereoscopic quality assessment metrics.

The most reliable, direct and dependable evaluation of S3D images and videos is subjective assessment, which is commonly performed by test participants who rate the perceived visual quality. Subjective analysis can also be done in other ways. For example, Malekmohamadi *et al.*[2] proposed an automatic, fast and accurate subjective quality estimation method based on decision trees for computing the subjective scores of S3D videos. However, the process of subjective analysis is expensive, cumbersome and extremely time-consuming, which makes it a rather inconvenient process to integrate in real-time systems. Therefore, these drawbacks provide motivation to develop efficient, fast and accurate objective S3D IQA/VQA metrics, the results of which are consistent with the subjective quality rating.

Objective S3D quality assessment techniques can be categorized into three types, depending on the availability of the distortion-free pristine or reference image/video contents. There are the full-reference (FR), reduced-reference (RR) and no-reference (NR) stereoscopic IQA/VQA metrics. FR methods use all the reference-quality image/video information as input to evaluate the quality of the distorted contents. RR methods use partial information of the reference-quality content as input for quality assessment, and NR methods are completely blind, which means that no information of the reference-quality content is available. As an evident consequence, FR methods generally have better performance, followed by RR and then NR methods. Yet in most practical scenarios, the distortion-free image/video content is not available for real-time use, which makes the development of NR IQA/VQA methods more important, compared to the other two methods.

This paper provides an exhaustive collection of various IQA and VQA techniques, including conventional statistical methods and machine learning algorithms as well. A total of 65 objective metrics is addressed, highlighting the specific S3D databases that are used for their performance evaluation. The remainder of the paper is organized as follows. Section 2 analyzes the different S3D IQA metrics and algorithms developed in the past years, subdivided depending on conventional and machine-learning-based approaches. Section 3 addresses S3D VQA techniques. A discussion on the topic is provided in Section 4. Section 5 concludes the paper.

2. S3D IMAGE QUALITY ASSESSMENT

This section highlights the scientific literature in the objective quality assessment of S3D images. Conventional techniques and machine-learning-based (deep-learning-based) techniques are addressed separately, and each part is further categorized into FR, RR and NR IQA methods, wherever applicable.

2.1 Conventional S3D IQA

2.1.1 FR objective metrics

The work of Shao *et al.*[3] explores an FR stereoscopic IQA method considering the binocular human visual characteristics and the cyclopean image concept. The stereoscopic images are categorized into non-corresponding binocular fusion and binocular suppression regions, after doing consistency check on the left and the right image and comparing matching error between the corresponding pixels in the binocular disparity calculation, and then each of these regions are independently evaluated by considering their binocular perception property. Finally, all the evaluated results are pooled into a final score. Experimental results on S3D images[4] show that the proposed metric displays great consistency with the subjective evaluation results.

Fezza *et al.*[5] proposed an FR IQA metric particularly for asymmetrically distorted S3D images. The work considers the binocular human visual characteristics by assigning weighting factors for the quality of the left and the right views, and based on the Binocular Just Noticeable Difference (BJND), the quality score of each region is modulated according to its perceptual significance. Experimental results performed on LIVE 3D Phase II database[6] show that the proposed metric outperforms the other metrics for both asymmetrically and symmetrically distorted images, especially for the latter.

A three-step FR IQA method by learning binocular receptive field properties is proposed by the work of Shao *et al.*[7] and the results of the proposed algorithm are recorded by using images from the NBU 3D IQA database, LIVE 3D IQA Phase I[8] database, LIVE 3D IQA Phase II database, the MICT stereoscopic image database[9] and the CML database.[10] In the training phase, the latent structure of the distortion-free reference image is captured. In the quality estimation phase, the sparse feature similarity (SFS) index and the global luminance similarity (GLS) index is calculated. Finally, in the last step, the sparse energy and the sparse complexity is calculated as the basis of the binocular combination with respect to the previously learnt dictionaries and the estimated sparse coefficient vectors.

Zhang *et al.*[11] proposed an FR IQA method which is a 3D most apparent distortion (MAD) algorithm (3D-MAD). It is based on binocular lightness and contrast perception, and operates in two stages. In the first stage, the left and the right views of the stereoscopic image is given as inputs to the conventional 2D MAD algorithm and the quality of the combined binocular view is estimated via a weighted sum of the two estimates. In the second stage, intermediate maps are generated corresponding to the pixel-based contrast and the lightness distance, and quality of the cyclopean view is evaluated by measuring the statistical-difference-based features obtained from the reference and the distorted S3D images. The final quality score is obtained by combining the estimates of the previous two stages. The algorithm provides great results when tested on the LIVE 3D image Phase I database, the MCL-3D image database[12] and the IRCCYN/IVC 3D image database.[13]

Li *et al.*[14] developed an FR IQA metric for S3D images based on the disparity-gradient-phase-congruency similarity (DGP-SIM). Measurements of the three feature similarity indexes of the extracted disparity maps of the distortion-free reference and distorted stereoscopic pairs, gradient magnitude maps and phase congruency maps of distortion-free reference and distorted cyclopean images, are combined together to obtain a single stereoscopic image quality index score. The results of the experiments performed on the stereopairs of the LIVE 3D IQA Phase I and Phase II databases show that it is highly consistent with subjective scores.

Md *et al.*[15] proposed an FR IQA algorithm called STeReoscopic Image Quality Evaluator (STRIQE) for predicting the quality of stereoscopic images using natural stereoscopic scene statistical (NSSS) models. In this method, generalized Gaussian density (GGD) fits of luminance wavelet coefficients are used along with the correlated relationship of luminance and disparity wavelet coefficients to successfully predict the quality score of the input S3D image. The efficiency of the proposed algorithm is verified using the LIVE 3D IQA Phase I, the LIVE 3D IQA Phase II and the IRCCYN databases.

2.1.2 RR objective metrics

Xu *et al.*[16] developed an RR IQA metric based on measuring structural degradation and the saliency-based parallax compensation model (SSPM). The algorithm works in three main stages of feature extraction of the reference and the distorted images, saliency-based parallax compensation and the effective nonlinear combination of the extracted features. It is shown that the SSPM metric outperforms other methods when evaluated on the S3D images of the LIVE 3D Phase I and Phase II databases.

The work of Ma *et al.*[17] introduces an RR S3D IQA metric based on the entropy of the gradient primitives. In this method, after computing the gradient maps of the left and the right views, the binocular perceptual information of gradient (GBPI) is shown by the distribution statistics of the visual primitives in the gradient maps of the left and right stereo image pairs, which is extracted by using sparse representation. The monocular cue is represented by the entropy of gradient maps of the stereo image pairs, and the binocular cue is represented by using their mutual information. The quality features are represented by the difference between the GBPIs of the distortion-free reference images and distorted images, and then a non-linear relationship is simulated between these features and human opinion, using the kernel ridge regression (KRR). The proposed metric outperforms other SIQA algorithms when evaluated over the LIVE 3D Phase II database.

Wan *et al.*[18] proposed an RR IQA metric based on sparse representation and natural scene statistics. The visual information is measured by using the distribution statistics of the classified visual primitives, which is extracted by using sparse representation, and the binocular fusion process is simulated by using the binocular cue, which is derived from the mutual information of the classified primitives between the left and the right views. The natural losses are computed by using the natural scene statistics of locally normalized luminance coefficients, and the quality of the S3D image is computed by the difference of the visual information and the natural scene statistics between the reference and the distorted images, as a prediction function trained using a support vector regression (SVR). The proposed metric shows outstanding performance when evaluated over the

LIVE 3D IQA (Phase I and Phase II) databases, NBU-MDSID Phase II database,[19] and the Waterloo IVC 3D (Phase I[20] and Phase II[21]) databases.

2.1.3 NR objective metrics

Sazzad *et al.*[9] proposed an NR SIQA model for both symmetric and asymmetric JPEG-coded S3D images by evaluating segmented local features, such as flat and texture areas, edges, blockiness and zero crossing rate of the block of images, for artefacts and disparity. Experiments conducted over 490 stereoscopic image pairs shows that the proposed model performs very well over a varied range of images and at varied levels of distortion.

The ODDM algorithm – the Ocular Dominance theory and Degree of parallax-based Distortion Metric – is a four-stage NR IQA metric proposed by Gu *et al.*,[22] which combines the concepts of a 2D NR IQA method[23] (the ocular dominance theory and degree of parallax). In this method, the ocular dominance difference is calculated for the left and the right 2D image qualities, then the compensating quantity from different degrees of parallax is calculated, and the 2D quality scores and the compensating quantity are combined while reducing ocular dominance difference to obtain the final quality score of the S3D image. The ODDM algorithm is very effective in predicting S3D image qualities and its high performance is shown over the Toyama database.[24]

Ryu *et al.*[25] proposed a top-down NR IQA metric, which models the binocular quality perception of the HVS by computing the perceptual blurriness and blockiness scores of the individual views of the stereo pair and then combines them to achieve the final quality index of the S3D image. Experiments over the Waterloo IVC Phase I and II, LIVE 3D IQA and DIML[26] databases show that the proposed NR model is highly consistent with the subjective analysis scores and performs better than most of the existing FR methods.

Su *et al.*[27] introduced an NR S3D IQA metric named Stereoscopic/3D BLind Image Naturalness Quality (S3D-BLINQ) Index and it is the first quality assessment algorithm to utilize both univariate and generalized bivariate along with correlation natural scene statistics (NSS) models to predict and quantify perceived distortion on stereoscopic image pairs. Experimental results over the LIVE 3D IQA Phase II database shows that the S3D-BLINQ model outperforms state-of-the-art FR and NR 3D IQA algorithms on both symmetrically and asymmetrically distorted stereo image pairs.

Shao *et al.*[28] proposed a two-stage NR stereoscopic IQA model for the prediction of S3D image quality based on monocular feature encoding and binocular feature combination. At the training stage, the multi-scale dictionaries from a given set of training data are trained using a sparse representation algorithm, while an ML-based trainer is constructed to model the process of the HVS activities by mapping the extracted feature vectors to the corresponding subjective difference mean opinion score (DMOS) value on different training data sample. At the testing stage, using the learnt regression model, the quality scores of the left and the right view images are predicted and then combined, depending on the binocular features. The performance of the model is validated over the NBU 3D IQA database and LIVE 3D IQA Phase I database.

Another two-stage blind NR IQA was proposed by Shao *et al.*[29] based on the learning of receptive fields (RFs) and the construction of quality lookups (QLs) to successfully replace human opinion scores without causing any performance loss. In the training phase, the local quality lookups (LQLs) and global quality lookups (GQLs) are constructed by learning local RFs (LRFs) and global RFs (GRFs) from the original and distorted images. In the testing phase, the optimal LRF and GRF indexes obtained from the learnt LQLs and GQLs are combined together to compute the final quality score. To demonstrate the high consistency of the proposed algorithm with the subjective evaluation score, experimental results have been tested over the NBU 3D IQA database, LIVE 3D IQA Phase I database and LIVE 3D IQA Phase II database.

A four-step NR IQA model was also proposed by Shao *et al.*,[30] which uses joint sparse representation and it is based on feature encoding, feature filtering, spatial pooling together and a quality prediction function, using SVR. The proposed metric simplifies the prediction of S3D image quality as a combination of feature-prior and feature-distribution, but since the metric still requires human opinion scores in order to train the regression model, hence it is not a “completely blind” model. The good performance of the metric in handling IQA issues of symmetrically and asymmetrically distorted images is validated using the NBU 3D IQA database, LIVE 3D IQA Phase I database, LIVE 3D IQA Phase II database, MICT stereoscopic image database, and CML database.

Khan *et al.*[31] proposed an NR model for IQA based on the joint statistical modeling of the wavelet subband coefficients of the stereoscopic image pair. In this method, the joint statistical texture features obtained using bivariate and multivariate generalized Gaussian distribution are added with the depth features extracted from the estimated disparity maps, and the final quality score of the stereoscopic image pair is predicted using the resultant features with the help of a machine-learning approach. The proposed method shows promising results when tested over the LIVE 3D IQA database.

Wang *et al.*[32] proposed an NR IQA based on visual saliency regions in the wavelet domain and wavelet transform. In this method, the visually significant regions are detected for the distorted image pairs, two separated cyclopean images created using the Gabor filtering and the structural-similarity-index-based (SSIM-based) stereoscopic algorithm and their corresponding depth maps, and then those images are segmented into respective patches. The phase amplitude and gradient features of the wavelet subband are obtained as the image features for the S3D image from a wavelet decomposition. The algorithm model is trained by SVR to find a mapping relationship between the features of S3D image quality and the DMOS, as well as to evaluate the objective perception quality score. Experimental results are verified using the LIVE 3D IQA Phase I and Phase II databases.

Appina *et al.*[33] proposed an NR S3D IQA algorithm called the Multi-Orient Naturalness Image Quality Evaluator (MO-NIQE). In this method, the steerable subband decomposition of the cyclopean image is performed at multiple orientations and the NIQE score and entropy score is computed from each subband. The computed quality scores of the steerable subbands are then pooled to obtain the overall perceptual quality score of the stereoscopic 3D image. The proposed algorithm shows outstanding performance on both symmetric and asymmetric distorted images when evaluated on the LIVE 3D IQA Phase I and Phase II databases.

Chen *et al.*[34] proposed an NR SIQA model which takes the stereo image pair as its input and successfully predicts its quality, as is verified using the LIVE 3D IQA database. In this method, a SSIM-based stereo algorithm is used to generate an estimated disparity map and a set of multi-scale Gabor filter responses are computed on the input by using a filter bank. The 2D features extracted from the cyclopean image generated using the inputs, estimated disparity map and the Gabor filter responses, along with 3D features extracted using the estimated disparity map and an uncertainty map created by the stereo matching algorithm, are then together put into a quality estimation module which gives the estimation of the quality of the input stereo image pair.

2.2 Machine-learning-based S3D IQA

2.2.1 NR objective metrics

Sun *et al.*[35] proposed two NR SIQA methods based on deep neural networks for learning local quality-aware structures of salient regions of S3D images and global score regression, named the one-column deep SIQA (ODSIQA) model and the three-column deep SIQA (TDSIQA) model. While both the one-column and the three-column convolutional neural network (CNN) model share the same implementation approach, they differ in their inputs. The one-column model takes only the cyclopean view as the input for learning local binocular features, whereas the three-column model takes the cyclopean as well as the left and the right views as its CNN inputs for imitating HVS perception, which is affected by both monocular and binocular properties, such as binocular features and rivalry. It is clear that the ODSIQA performs poorly on asymmetrically distorted images since it does not consider monocular properties and binocular rivalry, and that the TDSIQA shows high performance compared to all other methods, except Wang's FR SIQA metric[20] when evaluated over LIVE 3D IQA Phase I database, LIVE 3D IQA Phase II database and Waterloo IVC Phase II database.

Shen *et al.*[36] introduced an NR IQA method which takes image distortion, depth perception and binocular combination perception into consideration. In this method, based on the Gaussian average SSIM, a disparity search algorithm is designed, and rivalry maps, weight maps and cyclopean image are generated. Then, 2D image features of the left and the right views and the 3D image features of the cyclopean view are extracted for image distortion, disparity map features are extracted for depth perception, and the binocular rivalry features from the cyclopean image and the disparity map are extracted for binocular combination perception. Machine learning (ML) is used to map the extracted features to corresponding quality scores, and thus, the final 3D image quality score is predicted using a support vector machine (SVM) regressor. Experimental results over the LIVE 3D IQA

Phase I and Phase II database show that the proposed method outperforms current FR-SIQA and NR-SIQA metrics.

A blind deep quality evaluator (DQE) for 3D stereoscopic images, called 3D-DQE, was proposed by Shao *et al.*[37] based on monocular and binocular interactions. In this method, two individual 2D deep neural networks (2D-DNNs) are trained from monocular and cyclopean images in order to model the quality prediction process of monocular and binocular views, and then the estimated quality scores of the 2D monocular and the cyclopean view are combined together using different weighting schemes. Since the model needs human opinion scores to train the regressor on the dataset, hence, it is not a “completely blind” model. The high performance of the proposed model is validated over the NBU 3D IQA database, LIVE 3D IQA Phase I database, LIVE 3D IQA Phase II database and MCL-3D image quality database.

Li *et al.*[38] proposed a two-stage NR IQA for stereoscopic images based on using NSS along with a Deep Belief Network (DBN). In this method, after classifying the distorted image into symmetrical or asymmetrical distortions using the characteristics of wavelet domain and disparity information, where such wavelet domain features are classified into distortion types using a DBN, a mapping relationship is established between the NSS features and the stereoscopic image quality, depending upon the classified distortion type. The proposed algorithm delivered good performance against other 3D NR and FR IQA algorithms when tested over the LIVE 3D database.

Oh *et al.*[39] proposed a deep NR S3D image quality evaluator called DNR-S3DIQE based on local-to-global feature aggregation and deep CNNs. It takes a pair of S3D normalized images as input, without any reference to its depth, and combines the feature extraction and regression processes together. It automatically extracts any valuable local features and aggregates them into global features via local and global feature extraction and bi-directional update. Then, using data preprocessing, the local patches of S3D image pair are normalized, the local features from each of the patch pair are automatically extracted and local quality scores are computed. The final S3D image quality score is accurately predicted by aggregating the extracted local features into global features. The performance of the DNR-S3DIQE is verified using the LIVE 3D IQA Phase I and Phase II databases.

A blind SIQA is proposed by the work of Yang *et al.*[40] based on stacked auto-encoders (SAE) and a fusion of the cyclopean channel theory and the binocular summation/difference channel theory. Three SAEs are trained in an unsupervised manner for computing deep the quality-aware features from the cyclopean, summation and difference images. Then, two separate SVRs are trained: one for using the deep features of the cyclopean image and its corresponding subjective score, and another for using deep features of the summation image, the difference image and their subjective scores, respectively. Finally, by using a weighted sum, the output partial scores from the two SVRs are combined together to compute the total quality score of the S3D image. The proposed algorithm outperforms the state-of-the-art 3D IQA methods when tested over the LIVE 3D IQA Phase I database, the LIVE 3D IQA Phase II database, the Waterloo IVC 3D Phase I database and the Waterloo IVC 3D Phase II database.

Yang *et al.*[41] proposed another NR SIQA biologically-inspired metric, called Deep Edge and COlor Signal INtegrity Evaluator (DECOSINE) based on the perception route of the HVS from the eyes to the frontal lobe, while particularly focusing on the edge and the processing of color signals on the retinal ganglion cells and lateral geniculate nucleus. Since deep-learning-based SIQA methods need a long training time, S-SAE is used to model the structure of the visual cortex, which is a first of its kind. The DECOSINE algorithm computes the edge quality index and color quality index, and the final overall quality score of the stereoscopic image is calculated as a weighed sum of these two scores. The DECOSINE algorithm outperforms nine existing IQA metrics when evaluated over the LIVE 3D IQA Phase I database, the LIVE 3D IQA Phase II database, the Waterloo IVC SIQA Phase I database, the Waterloo IVC SIQA Phase II database and the IVC SIQA database.

Ding *et al.*[42] proposed an NR SIQA method using CaffeNet for adaptive quality-aware monocular feature extraction to classify images with respect to their perceptual quality. The CNN model is trained using distorted images from the LIVE 2D IQA database[43] and distorted stereoscopic image pairs from the LIVE 3D IQA Phase I database. Then, the captured monocular features are fused together using visual saliency models and binocular disparity maps are used to get the multi-scale statistical features. Finally, the objective quality score of the S3D image is derived by synthesizing the fused CNN features and the disparity features using an SVR. The LIVE 3D

IQA Phase I database and the LIVE 3D IQA Phase II database is used to verify the high performance of the proposed method.

A three-stage NR IQA model is proposed by the contribution of Messai *et al.*[44] based on the HVS perception model and using four CNNs as prediction models. The presence of binocular rivalry/suppression is considered to overcome the asymmetric distortion issue in stereoscopic images during the formation of the cyclopean image, and this image is then divided into four patches. The four trained CNN models are then used to calculate the quality scores of the four cyclopean patches, and the final quality score of the stereoscopic image is found by averaging the outputs of the four CNNs. The excellent performance of the proposed model is evaluated over the LIVE 3D IQA Phase I and Phase II databases.

Based on the hierarchical dual-stream interactive nature of the HVS perception model, an end-to-end dual-stream interactive stereoscopic image quality assessment network, called StereoQA-Net, is proposed by the work of Zhou *et al.*[45] as an NR SIQA method. The StereoQA-Net contains two primary sub-networks in multiple convolutional layers for the left and the right views, and it integrates both of these in accordance with the fusion and disparity information of the HVS, by performing summation and subtraction of the corresponding feature maps for the distorted patch pairs. The performance of the method is validated using the LIVE 3D IQA Phase I and Phase II databases.

Kim *et al.*[46] proposed a blind IQA method for predicting the quality of stereoscopic images using pooling of the patch features to the image features. In this method, patch-based CNNs are used to overcome the shortage of training data and every stereoscopic image is divided into patch pairs. Then, the patch features extracted from each patch pair using the CNNs are automatically pooled into image features. Finally, by using mean opinion score (MOS), the model parameters of the trained CNN are iteratively updated. The LIVE 3D IQA Phase I database is used to verify the better performance of the proposed method compared to other NR S3D IQA and FR S3D IQA algorithms.

An NR IQA method for S3D images is proposed by the work of Xu *et al.*[47] based on the concept of transfer learning and saliency-guided feature consolidation and feature extraction through fine-tuning strategy using CNNs. Both CaffeNet and GoogLeNet are adopted for this fine-tuning strategy. The fine-tuned CNN model is then used to extract the quality-aware features of the left and the right views by using the transfer learning concept, and these features are then combined as a linear weighted fusion, where the weights of each image is found from its saliency maps. The additional features considered are the statistical characteristics of the disparity map in a multi-scale manner. Finally, SVR is used to obtain the objective score for each stereoscopic image pair. The proposed method outperforms many NR and FR IQA methods, which is validated using the LIVE 3D IQA Phase I and Phase II databases.

Zhao *et al.*[48] proposed a multi-scale dilation CNN (MSDCNN) framework for NR SIQA, by using dilation convolution neural networks (DCNNs) to imitate the multi-scale characteristics of information processed by the human brain. Caffe is used to train the MSDCNN and it takes a color cyclopean image split into 40x40 small patches as its input. The entire body of the proposed framework is comprised of three multi-scale units cascaded together: a standard convolution layer, a global pooling layer and a fully-connected (FC) layer, which has one input and one output. Each convolution layer of the model has 3x3 kernels and the activation function used after the convolution layer is the rectified linear unit (ReLU). The stochastic gradient descent (SGD) method is used to train the MSDCNN model, and network parameters are adjusted in accordance with the Euclidean loss. The MSDCNN can accurately evaluate the quality score for distorted S3D images, which is shown using the LIVE 3D IQA Phase I and Phase II databases.

Liu *et al.*[49] proposed a two-stage blind IQA model based on hierarchical learning, which automatically predicts the perceived quality of S3D images. The model is named CAP-3DIQA, since it is used for the classification and prediction of 3D image quality. In the classification stage, images with the same types of distortions are grouped into the same subset and there are several such subsets depending on the type of distortions available. In the prediction stage, an image quality predictor having five different perceptual channels and taking the classified distorted image subsets as inputs is used to predict the quality scores of the images individually. Finally, the outputs of the five channels are combined together and fed into the regression module of the SVM to successfully predict the final quality score of the stereoscopic image. Experimental results verify the proposed method over the LIVE 3D IQA Phase I database, the LIVE 3D IQA Phase II database, and the MCL 3D image quality database.

Li *et al.*[50] proposed an NR S3D IQA metric, which is a double-channel CNN model with multiple-level fusion network (MLFNet) to achieve long-term feature fusion process, and can obtain feature extraction, fusion and processing simultaneously. The MLFNet is comprised of five CNN layers having ReLU activation function and a pooling layer following the first two convolutional layers, five concatenation (concat) layers having five squeeze-and-excitation (SE) blocks with five uniform modules embedded after each block, two global pooling layers, and three fully-connected layers, and it is simultaneously able to extract low-level and high-level features. It takes the left and the right view images as the input, and the final quality score of the stereoscopic image is available as output of the final FC layer. The proposed method showed excellent performance on symmetrically- and asymmetrically-distorted images when evaluated on the LIVE 3D IQA Phase I and Phase II databases.

Sim *et al.*[51] proposed a blind NR SIQA framework based on a binocular semantic channel trained on the ImageNet dataset for extracting semantic features and a binocular quality channel trained on the SIQA dataset to extract quality-aware features from the stereoscopic image. The binocular semantic channel consists of two deep convolutional neural networks (DCNNs) of the VGGNet structure in parallel which takes the left and the right views as its input, and uses the extracted semantic features as well as MOS/DMOS to obtain a mapping function score. The binocular quality channel also takes the left view and right images as its input and after the iterative task of low-pass filtering and down-sampling, local normalization of the input image pair, it uses the extracted quality-aware features and MOS/DMOS to learn a mapping relation between them and obtain a mapping function score. Finally, the perceptual quality of the stereoscopic image is obtained by combining both of these mapping function scores. Experimental results show that the proposed method demonstrates high consistency with subjective analysis scores and this is validated using the LIVE 3D IQA (Phase I and Phase II) databases and the Waterloo IVC (Phase I and Phase II) databases.

3. S3D VIDEO QUALITY ASSESSMENT

This section highlights the scientific literature in the objective quality assessment of S3D videos. Conventional techniques and machine-learning-based (deep-learning-based) techniques are addressed separately, and each part is further categorized into FR, RR and NR IQA methods, wherever applicable.

3.1 Conventional S3D VQA

3.1.1 FR objective metrics

Galkandage *et al.*[52] proposed a three-stage FR IQA metric based on the HVS model, including the temporal domain for the first time, which was extended to be used for VQA purposes by the optimized temporal pooling strategy. In this method, first a HVS model is develope,d which takes into consideration binocular suppression and recurrent excitation to provide better indication of the perception of depth, then a relationship between the binocular signals and subjective analysis scores is obtained by using a statistical analysis technique, and finally, the method is extended to be used for VQA by temporal pooling to provide two video quality metrics. The performance of these metrics shows high consistency between the subjective and objective scores, and it is evaluated through four image datasets and two video datasets. The only limitation of the method is that the stereoscopic input should not contain any vertical parallax.

An FR stereoscopic VQA algorithm called depth- and motion-based 3D video quality evaluator ($DeMo_{3D}$) is proposed by the work of Appina *et al.*[53] to estimate the perceptual quality of natural S3D videos based on the directional dependency between the depth and motion subband coefficients of the video frames, and this joint statistic is found using the bivariate generalized Gaussian distribution (BGGD) model. Since the coherence of the BGGD covariance matrix proportionally varies with the perceptual quality of video, the directional dependency between the components of depth and motion is measured using the coherence scores of the eigen values of this matrix. To compute the spatial quality score, a 2D FR IQA is used on the individual left and right views, and then the average of their frame-wise score is taken. Finally, the quality of the given natural stereoscopic video is determined by pooling the estimated coherence and spatial scores. The results indicating the high efficiency of the algorithm is verified over the IRCCYN,[54] the Waterloo IVC Phase I, and the LFOVIA[55] S3D video databases.

Md *et al.*[56] proposed an FR stereoscopic video quality assessment (FRSVQA) algorithm based on the spatial, depth and temporal qualities of S3D video on a per-frame basis. The spatial quality is estimated by using a spatial distortion map on each frame of the video, the depth quality is estimated by refining the spatial distortion map by using depth salient maps and the temporal quality is estimated by refining the spatial distortion map by using the inter-frame difference map at the places identified by motion edges. Finally, all the three estimated qualities are combined and their average over the frames provides the overall quality metric of the input S3D video. The proposed algorithm shows very good performance when tested over the IRCCYN S3D VQA database.

Galkandage *et al.*[57] proposed another FR VQA metric based on the motion sensitivity response of the complex cells of the HVS model in the primary visual cortex of the human brain. The output of these complex cells is modulated depending on local motion parameters, such as direction and amplitude. The main application of this method is that it helps in defining the binocular energy terms, that is, non-motion-sensitive and motion-sensitive energy terms, using the non-motion-sensitive and motion-sensitive characteristics of the S3D video on a frame-wise basis to mimic the working of the HVS model. These energy terms are successfully determined using a two-step multi-variate stepwise regression algorithm and experimental results are validated over the ROMEO project dataset,[58] the NAMA3DS1-COSPAD1 dataset and the Waterloo 3D-VQA[59] dataset.

A four-stage FR VQA algorithm called $FLOSIM_{3D}$ is proposed by the work of Appina *et al.*[60] based on the temporal, spatial and depth features of natural stereoscopic videos. In this method, the temporal quality features are extracted from both the left and the right views by an existing motion-based 2D stereoscopic VQA metric, the spatial quality features are estimated by using an existing 2D stereoscopic IQA metric for both left and the right views and the depth quality features are estimated with the help of depth maps for each frame of both left and right views. Finally, frame-level scores are obtained for the left and the right views by integrating these three estimated quality features, and the overall perceptual quality of the S3D video is computed by averaging the two view-wise scores obtained by pooling these frame-level scores. The proposed algorithm shows state-of-the-art performance over the IRCCYN dataset.

Silva *et al.*[58] designed an FR VQA metric named Stereoscopic Structural Distortion (StSD) for predicting the perceptual quality of compressed S3D videos based on the measurement of structural distortion, the measurement of asymmetric blur and the measurement of content complexity features, such as spatial and temporal features. The performance of the proposed metric is verified using 14 stereoscopic video samples and it is found to be highly consistent with the subjective analysis scores.

Hong *et al.*[61] proposed an FR VQA called 3D Perceptual Quality Index (3-D-PQI) to measure the perceptual quality of compressed stereoscopic videos based on the spatial distortion, temporal distortion and binocular viewing property of the HVS. In this method, the left view and the right views are separately processed, the spatial and temporal objective distortions are calculated, this objective noise is turned into perceptual noise by modelling two important HVS properties of contrast masking and motion masking as visual distortion sensitivity parameters, the local spatial and temporal distortions are pooled based on stereo visual saliency to accumulate local channel-separate noise, final perceptual quality index (PQI) of each frame in the left and the right views is calculated separately by fusing the measured spatial and temporal distortion, and finally, overall PQI of the S3D video is estimated by combining the PQI of the left and the right views by binocular fusion. Experimental results are verified using the NAMA3DS1-COSPAD1[62] database, the SVQA[63] database, and the Waterloo-IVC 3D video Phase I and Phase II database.[59]

3.1.2 RR objective metrics

Malekmohamadi *et al.*[64] proposed an RR VQA metric based on spatial pixel-wise neighbouring information and edge information of stereoscopic videos. For spatial neighbouring information, contrast features from grey-level co-occurrence matrices (GLCM) are measured for depth and color information for every frame of the original video and compressed video. For edge information, side information from the frames of the original video is extracted. Finally, the contrast measures from the GLCM as well as the edge information is sent via an auxiliary channel of very low bandwidth. The proposed metric also takes into consideration the color-to-depth ratio, since the color and depth views have different weights, which can help to enhance the performance of the proposed metric to accurately match with subjective scores for some particular values. Experimental results over a varied range of stereoscopic video samples show that the proposed metric has an average correlation of 0.82 to the subjective score when the color-to-depth ratio is approximately 4.

Appuhami *et al.*[65] proposed an RR VQA metric called Correlation Coefficient (CC) based on 3D structural tensors since the HVS model is highly sensitive to any structural information present in the view. In this algorithm, the salient points in the spatial domain and the temporal domain are selected separately for the original reference video, right view and left view, and the predominant energy distribution of these selected salient points is calculated by using the 3D structure tensor. Then, the estimated structure tensor values over all the selected salient points are averaged in order to find the quality indicator of the video frame. Finally, the estimated frame-level quality indicator scores over all the frames in the S3D video sequence is averaged to determine the final predicted quality indicator score. The results showing the outstanding performance of the model is verified by using the IRCCYN NAMA3DS1 database53 and the Kingston University database.[66,67]

Yu *et al.*[68] proposed a three-stage RR VQA method based on the binocular perception of the HVS model on the temporal characteristics of S3D video scenes. In this method, RR frame pairs are extracted from the input stereoscopic video by using motion intensity, which is defined based on the temporal characteristics present in the video. The extracted RR frame pairs are then divided into the binocular fusion portion (BFP) and the binocular rivalry portion (BRP). The cyclopean view is constructed using the BFP. This cyclopean view and the BRP are used to extract the GGD features. Quality indicators are computed for the BFP and the BRP, and a comparison between the original and distorted frame quality indicators is performed. Finally, the overall S3D video quality score is obtained by pooling these estimated quality indicators in the spatial and temporal domain. Experimental results are verified using the NAMA3DS1-COSPAD1 video database.

3.1.3 NR objective metrics

Sazzad *et al.*[69] proposed an NR continuous VQA method based on spatio-temporal segmentation for predicting the perceptual quality of S3D videos coded with MPEG-2 MP@ML with different bit rates. The left and the right views of the given S3D video is individually converted into frames and the partition of videos in the sub-temporal and temporal segment is performed using temporal segmentation with both the segments being partially overlapping. A segmentation algorithm is used to identify the edge and non-edge area of the frame content for spatial segmentation. Finally, artefact measure, disparity measure and temporal features measure are calculated separately for each temporal segment. It is found that the proposed method is highly adept at predicting the quality of a given stereoscopic video, which is verified by using subjective scores on various symmetrically- and asymmetrically-distorted videos.

An NR 3D VQ metric called NR-3VQM is proposed by the work of Solh *et al.*[70] using depth image-based rendering (DIBR). In this method, after deriving an ideal depth estimate for every pixel value, this estimate is then utilized to compute the three distortion measures, namely, temporal outliers (TO), temporal inconsistencies (TI) and spatial outliers (SO). By combining these three measures, the proposed metric NR-3VQM is determined, and the final quality measure of the stereoscopic video is estimated as the mean of the NR-3VQM matrix values. The metric is found to be highly consistent with subjective scores and its performance can parallel that of an FR VQA algorithm.

Ha *et al.*[71] proposed an NR stereoscopic video quality perception model (SV-QPM) for perceptual quality measurement based on the estimated disparity information in the absence of depth information of a given S3D video. To design the proposed model with the required visual features, temporal variance due to motion activity in the video, intra-frame disparity variation due to the difference in the disparity values between the neighbouring blocks in the spatial domain, inter-frame disparity variation due to disparity changes in the temporal direction and disparity distribution of frame boundary areas are all combined together by using a linear regression model. The proposed model shows highly promising results with a Pearson correlation coefficient (PCC) value of 0.808.

Han *et al.*[72] proposed an NR VQ Metric (NVQM), which is a modified version of the 2D Video Quality Metric ITU-T G.1070,[73] for the perceptual quality assessment of real-time S3D videos, while taking into consideration the left and the right image views, the network packet loss and the perceptual quality of videos with different bit rates. The left and the right views are processed separately and then combined at the display side to give the perception of depth, which is vital for 3D experience. Experimental results show that the proposed metric outperforms many of the existing objective metrics in quality evaluation.

Han *et al.*[74] proposed another extended NR objective 3D Video Quality Metric (eNVQM), which establishes a relation between the network packet loss and the quality of an S3D video, and it is based upon the bitrate of the video, packet loss rate and frame rate of the video. eNVQM shows better performance in predicting perceptual quality when compared to the SSIM and the video quality metric (VQM).

Hasan *et al.*[75] proposed an NR VQA metric based on a disparity index measured by feature dissimilarity information of the videos and the perceptual difference index depending on the detected edges. These two extracted features are then combined using the suppression of the binocular vision concept to provide the final video quality assessment metric. The accuracy of the proposed metric is verified using video sequences from the RMIT3DV[76] and EPFL[77] database.

Mahmood *et al.*[78] proposed an NR VQA metric called MD-QA based on the weighted values of motion vector features and depth map features of S3D videos. The depth map information is extracted by generating the disparity map for the left and the right views of the stereoscopic video, and the motion quality factor is extracted by generating motion vector map for every frame of the video. Finally, the weighted values for both depth and motion quality factors are pooled using a non-linear regression function to provide the final overall quality score for the video. The performance of the proposed metric is tested over the EPFL 3D stereoscopic video database.[77, 79]

Appina *et al.*[80] proposed a supervised NR natural stereoscopic VQA algorithm called Video QUality Evaluation using MOtion and DEpth Statistics (VQUEMODES) based on modelling the joint statistical dependency of the subband coefficients of motion and depth statistics of the S3D video using BGGD. The motion quality feature and depth quality feature are represented by using the BGGD model parameters, which are estimated at every subband on a frame-wise basis, and the spatial quality feature is represented by applying the 2D NR IQA model (NIQE)[81] on a frame-wise basis for the left and the right views. Then, all the frame-wise estimated features are consolidated along with the frame-wise DMOS score label for supervised learning purposes by using SVR. Finally, the frame-wise estimated quality scores are averaged to provide the overall perceptual quality prediction for the given S3D video. The success of the proposed algorithm is verified using the IRCCYN and the LFOVIA S3D databases.

Another NR VQA algorithm called Motion and Disparity-based 3D video quality evaluator ($MoDi_{3D}$) is proposed by the work of Appina *et al.*[82] based on modelling the joint statistical dependency of the subband coefficients of motion and disparity statistics of the S3D video by using BGGD, along with developing a new S3D video dataset called LFOVIAS3DPh2 S3D, which contains videos with H.264 compression, H.265 compression, blur and frame freeze distortions, along with the distortion-free pristine videos. In this method, motion vectors and disparity maps are estimated on a frame-wise basis and then modelled using BGGD parameters, and spatial quality features are estimated using a 2D NR IQA model (NIQE) for both the left and the right views. Finally, by pooling the estimated likelihood that the MVG model parameters of the test video is coming from the MVG model of the pristine video, the estimated global motion strength found by performing the average of the SSIM scores of successive video frame scores and the estimated spatial feature quality score, we can determine the overall perceptual quality score of the stereoscopic video. Experiments over the IRCCYN dataset, the Waterloo IVC Phase I dataset, the LFOVIA dataset and the proposed LFOVIAS3DPh2 S3D dataset show that the proposed algorithm results are highly consistent with the subjective scores and performs better than many other VQA methods.

Yang *et al.*[83] developed two stereoscopic 3D video databases called the TJU-SVQA Phase I and the TJU-SVQA Phase II, and proposed an NR S3D VQA model based on the joint contribution of the spatial, temporal and spatio-temporal domains. In this method, the left view and the right view frames are separately processed, binocular summation/difference in the spatial domain and the inter-frame cross maps in the spatio-temporal domain are produced and their quality perception features are extracted. Then, in the temporal domain, the optical flow features are extracted to estimate the degree of distortion. Finally, the quality scores from each domain are acquired using sparse representation, dictionary learning and SVR, and these scores are then pooled together to determine the overall video quality. The results are verified over the TJU-SVQA Phase I database, the TJU-SVQA Phase II database and the NAMA3DS1-COSPAD1 database.

Yang *et al.*[84] proposed another NR SVQA algorithm based on the perception of motion by the HVS. In this method, the left and the right views of the video are separately processed and the keyframe sequences from both videos are selected based on the concept of motion, masking of human binocular vision. Then, based on stereo perception, these selected keyframes are encoded and binocular summation and difference operations are performed on them. By using local binary patterns from three orthogonal planes (LBP-TOP), the texture features and motion information are obtained from the difference and the summation maps. Finally, by using a SVM, the overall quality of the S3D video is obtained as its output. Experimental results are verified over the NAMA3DS1-COSPAD1 dataset and the QI-SVAQ dataset.

Hou *et al.*[85] developed an NR S3D VQA algorithm called Stereoscopic Video Integrity Predictor using OLGF Statistics (SVIPOS) based on a proposed oriented local gravitational force (OLGF) descriptor in the space-time domain, which is an extension of an existing local gravitational force descriptor with two added new components of relative local gravitational force magnitude and orientation. Considering the left and the right views of the video sequences as input parameters, the cyclopean image and product image is generated in the spatial domain to measure the correlation between these two videos, and considering only the left video sequence, a frame difference image is generated in the temporal domain. The local gravitational force responses for the generated cyclopean image, product image and the frame difference image are computed by using the proposed OLGF model and statistical features are extracted by using these three computed responses. Finally, these extracted features are mapped to stereoscopic video quality predictions by using an SVR, in order to determine the overall perceptual quality of the S3D video. The efficiency of the algorithm is proven by using the NAMA3DS1-COSPAD1 database, the SVQA database[63] and the Waterloo IVC Phase I database.

3.2 Machine-learning-based S3D IQA

3.2.1 FR objective metrics

Xu *et al.*[86] proposed an FR method called Convolutional Neural Network with 3D kernels (C3D) for VQA (C3DVQA) based on combining the feature learning and score pooling process into a single spatio-temporal feature learning process. The spatial features are extracted by using two 2D convolution layers on the distorted frames and the residual frames which are the difference between the reference frames and the distorted frames, and then these features are concatenated to show the spatio-temporal context of the video. The spatio-temporal features are then extracted by using four 3D convolutional layers since they are able to obtain the temporal masking effects in the video frames. Finally, by using a global average pooling layer to represent the amount of perceived distortion, followed by two fully connected layers to represent the nonlinear relationship between the subjective evaluation score and the perceived distortion, the overall quality of the S3D video is obtained. Experimental results are verified by using the LIVE 3D video dataset[87] and the CSIQ video dataset.[88]

Reddy *et al.*[89] proposed an FR VQA framework called Deep Video QUality Evaluator (DeepVQUE) based on deep 3D convolutional neural network (3D ConvNet) models for the extraction of spatio-temporal features of the videos with respect to their distortion-free versions. These spatio-temporal feature qualities are estimated by using distance measures like the l1 or l2 norm to the volume-wise pristine videos and the distorted 3D ConvNet features, and a common FR image quality assessment (FRIQA) method is utilized to estimate the spatial quality features. Finally, SVR is applied to the estimated spatial and spatio-temporal quality features to obtain the overall objective quality score of the stereoscopic video. The efficient performance of the proposed algorithm is validated over the LIVE 3D database, the EPLE PoliMI database[90] and the LIVE Mobile database.[91]

3.2.2 NR objective metrics

Yang *et al.*[92] proposed an NR VQA model based on saliency maps and sparse representation. In this method, after extracting images from the left and the right views, the left view frames and right view frames are combined together and the sum maps of the 3D saliency maps are used to gather information about their spatio-temporal domain, saliency and background. Then, the sum map of the 3D saliency is decomposed into coefficients by using sparse representation and features extracted by using these coefficients are fed into an SAE. Finally, the output of the SAE is put as input to the SVM and the output of the SVM provides the overall quality score of the stereoscopic video. The performance of the proposed model in accurately predicting video quality score is validated over the NAMA3DS1-COSPAD1 database and the QI-SVQA video database.[63]

Yang *et al.*[93] proposed another NR S3D VQA model based on 3D CNN to extract a large amount of local spatio-temporal information as well as global temporal information by dividing the stereoscopic video into many cubic patches and feeding each cubic patch as an input to the 3D CNN. Average pooling is done in the spatial dimension to integrate the predicted quality score for each input cubic patch, and the overall quality score of the stereoscopic video is estimated using a quality score fusion strategy while considering global temporal information by computing the weight of each patch segment in the temporal dimension on the basis of motion intensity. The efficient performance of the proposed method is verified by using the NAMA3DS1-COSPAD1 database and the QI-SVQA database.[63]

Zhou *et al.*[94] proposed an NR SVQA method named End-to-end Dual stream deep Neural network (EDN) based on incorporating two subnetworks of two CNNs, having identical configuration and shared parameters for processing the left and the right views separately. The architecture of these subnetworks is designed based on AlexNet,[95] and it takes the left and the right view image patches as its input to estimate their perceptual quality. Then, convolutional layers are used to decrease the feature map sizes by using convolutional kernels via max pooling, and a fusion layer is utilized at the output end of the sub-networks to fuse the synthesized left and right views in order to provide depth perception. By using FC layers, regression is performed for both the left and the right view input patches to obtain a single quality score. Finally, the estimated quality score is verified using the ground truth quality score by feeding it into a Euclidean loss function. Results for the proposed method is verified using the NAMA3DS1-COSPAD1 dataset.

Varga *et al.*[96] proposed an NR VQA algorithm for predicting the quality of stereoscopic videos using frame-level deep features, which are extracted from successive resized video frames with cropped centre patches by using a pretrained CNN along with a long short-term memory (LSTM) network. In this method, the video sequence is considered as a time series of deep features extracted by using the CNN and the long-term dependencies are learnt with the help of the LSTM network. The CNNs can be pretrained in three different ways, namely, AlexNet,[94] Inception-V3[97] and Inception-ResNet-V2,[98] depending on the resizing and centre-cropping of consecutive video frame patches. After the LSTM receives these input video data sequence, they are run through the pretrained CNN and the frame-level feature vector is acquired in the form of a matrix after the removal of the last softmax as well as the last FC layer. Finally, the perceptual quality of the stereoscopic video is obtained by transferring this feature matrix as an input to the LSTM network. The proposed algorithm is shown to outperform other algorithms when tested using the Konstanz natural video quality (KoNViD-1k) database[99] and the LIVE VQA database.

Another NR VQA method is proposed by the work of Varga[100] based on frame-level deep feature vectors extracted by using pretrained CNNs (Inception-V3 and Inception-ResNet-V2). These CNNs are fine-tuned via the transfer learning process, they accept the resized and centre-cropped patches of the video frames as input and they extract the required frame-level deep features as their output. Then, temporal pooling is performed on an element-by-element basis for each frame-level feature vector to create one single video-level feature vector for each of the video sequence. Finally, a trained SVR is used to map the video-level feature vector to a video quality score. The KoNViD-1k and the LIVE 3D VQA database is used to verify the experimental results.

Feng *et al.*[101] proposed an end-to-end NR stereoscopic VQA model by using a multi-stage growing attention (MSGA) strategy at the fusion channel of the proposed network to extract spatio-temporal features, which is a first of its kind. The model comprises of the left channel, the right channel and the MSGA fusion channel, along with 3D convolution layers and FC layers. It takes the left and the right blocks of the video as input and extracts the spatio-temporal features by using 3D convolution. The final output of the proposed model provides the predicted video quality score, which is found to be highly consistent with HVS perception, as shown by using the NAMA3DS1-COSPAD1 and the QI-SVQA database.

4. DISCUSSION

The primary aim of all objective S3D quality metrics is to predict the perceptual quality of stereoscopic images and videos as accurately as possible, such that their predicted results are highly consistent with the subjective assessment results. This is a highly challenging issue in 3D signal processing, since the different types of quality degradation directly affects the QoE for a viewer, and hence, the development of novel, efficient, non-cumbersome and accurate S3D IQA/VQA metrics is extremely important. At the time of this paper, 3D applications have reached all facets of technology, such as television, mobile phones, gaming, movies, medical devices and many more, which leads to rising demands for different types of quality assessment metric to meet the dire needs. For example, Dehkordi *et al.*[102] proposed an FR VQA metric particularly for measuring the perceptual quality of 3D videos of mobile, and it is based on the quality of the monocular left and right views, the quality of the cyclopean view and the quality of depth maps. In this method, the effect of video resolution, viewing distance and the dimensions of the screen on the perceptual quality is also taken into consideration. The proposed method is found to be 82 percent consistent with subjective analysis scores, which is better than existing metrics for the same purpose.

Stereoscopic 3D images and videos are widely used by the gaming industry to provide a realistic visual experience.[103] The demands related to the high levels of user experience led to significant advancements in the development of suitable objective metrics. Even though a lot of efforts are being made to provide near-immersive experience to the users of video games,[104–106] more research is needed to overcome the side effects of 3D gaming, such as nausea and motion sickness.[107, 108] With the rapid advancements in medical technologies, there is a similarly high demand for stereoscopic medical imaging devices.[109] In a number of published works,[109–113] we can clearly see that quite a few of the works reviewed in this paper are utilized in medical contexts. Stereo 3D videos is also present in the designing of Automated driving systems (ADSs)[114, 115] and automatic parking systems[116] for providing visual depth perception. Standard IQA and VQA methods plays a vital role in monitoring Quality of Service (QoS) by different networks providers of stereoscopic images/videos, since transmission over wired/wireless channels can lead to different forms of degradation and the general loss of the stereo content. The most direct application for objective quality metrics is for monitoring and benchmarking different image and video processing datasets, algorithms and systems to maintain a certain uniform quality control.

5. CONCLUSION

Objective quality metrics provide an economical, efficient and effortless method for the assessment of the quality of stereoscopic images and videos, thus, avoiding the need for the laborious task of subjective quality assessment. In this paper, we provided a detailed analysis of the different objective quality metrics for S3D images and videos. Our work categorizes the metrics by the availability of the distortion-free, reference-quality content – such as full-reference, reduced-reference and no-reference metrics – and they are also differentiated by the utilization of machine learning. The discussion on the practical applications of such objective metrics particularly highlights their relevance and importance in different usage contexts. Additionally, as the paper covers all the various S3D databases used for evaluation, this may support the decision making processes of related future research efforts. We conclude that machine-learning-based S3D IQA and VQA metrics are continuously and quite rapidly emerging in the scientific literature, and their presence within practical usage contexts is expected to become more dominant in the upcoming years.

ACKNOWLEDGMENTS

The scientific efforts leading to the results reported in this paper were 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 the National Research Development and Innovation Fund based on the charter of bolster issued by the National Research Development and Innovation Office under the auspices of the Ministry for Innovation and Technology, Hungary.

REFERENCES

[1] Blake, R., Westendorf, D. H., and Overton, R., "What is suppressed during binocular rivalry?," *Perception* **9**(2), 223–231 (1980).

[2] Malekmohamadi, H., "Automatic subjective quality estimation of 3D stereoscopic videos: NR-RR approach," in [*3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*], 1–4, IEEE (2017).

[3] Shao, F., Lin, W., Gu, S., Jiang, G., and Srikanthan, T., "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Transactions on Image Processing* **22**(5), 1940–1953 (2013).

[4] Zhou, J., Jiang, G., Mao, X., Yu, M., Shao, F., Peng, Z., and Zhang, Y., "Subjective quality analyses of stereoscopic images in 3DTV system," in [*Visual Communications and Image Processing (VCIP)*], 1–4, IEEE (2011).

[5] Fezza, S. A., Larabi, M.-C., and Faraoun, K. M., "Stereoscopic image quality metric based on local entropy and binocular just noticeable difference," in [*International Conference on Image Processing (ICIP)*], 2002–2006, IEEE (2014).

[6] Chen, M.-J., Cormack, L. K., and Bovik, A. C., "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing* **22**(9), 3379–3391 (2013).

[7] Shao, F., Li, K., Lin, W., Jiang, G., Yu, M., and Dai, Q., "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Transactions on Image Processing* **24**(10), 2971–2983 (2015).

[8] Moorthy, A. K., Su, C.-C., Mittal, A., and Bovik, A. C., "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication* **28**(8), 870–883 (2013).

[9] Sazzad, Z. M. P., Yamanaka, S., Kawayokeita, Y., and Horita, Y., "Stereoscopic image quality prediction," in [*International Workshop on Quality of Multimedia Experience*], 180–185, IEEE (2009).

[10] Lin, Y.-H. and Wu, J.-L., "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Transactions on Image Processing* **23**(4), 1527–1542 (2014).

[11] Zhang, Y. and Chandler, D. M., "3D-MAD: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception," *IEEE Transactions on Image Processing* **24**(11), 3810–3825 (2015).

[12] Song, R., Ko, H., and Kuo, C., "MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source," *arXiv preprint arXiv:1405.1403* (2014).

[13] Benoit, A., Le Callet, P., Campisi, P., and Cousseau, R., "Quality assessment of stereoscopic images," *EURASIP Journal on Image and Video Processing* **2008**, 1–13 (2009).

[14] Li, F., Shen, L., Wu, D., and Fang, R., "Full-reference quality assessment of stereoscopic images using disparity-gradient-phase similarity," in [*China Summit and International Conference on Signal and Information Processing (ChinaSIP)*], 658–662, IEEE (2015).

[15] Md, S. K., Appina, B., and Channappayya, S. S., "Full-reference stereo image quality assessment using natural stereo scene statistics," *IEEE Signal Processing Letters* **22**(11), 1985–1989 (2015).

[16] Xu, Q., Zhai, G., Liu, M., and Gu, K., "Using structural degradation and parallax for reduced-reference quality assessment of 3D images," in [*International Symposium on Broadband Multimedia Systems and Broadcasting*], 1–6, IEEE (2014).

[17] Ma, J., Zhao, X., and Xu, Y., "Reduced-Reference Stereoscopic Image Quality Assessment Based on Entropy of Gradient Primitives," in [*5th International Conference on Signal and Image Processing (ICSIP)*], 206–209, IEEE (2020).

[18] Wan, Z., Gu, K., and Zhao, D., "Reduced reference stereoscopic image quality assessment using sparse representation and natural scene statistics," *IEEE Transactions on Multimedia* **22**(8), 2024–2037 (2019).

[19] Shao, F., Gao, Y., Jiang, Q., Jiang, G., and Ho, Y.-S., "Multistage pooling for blind quality prediction of asymmetric multiply-distorted stereoscopic images," *IEEE Transactions on Multimedia* **20**(10), 2605–2619 (2018).

[20] Wang, J., Rehman, A., Zeng, K., Wang, S., and Wang, Z., "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Transactions on Image Processing* **24**(11), 3400–3414 (2015).

[21] Wang, J., Zeng, K., and Wang, Z., "Quality prediction of asymmetrically distorted stereoscopic images from single views," in [*International Conference on Multimedia and Expo (ICME)*], 1–6, IEEE (2014).

[22] Gu, K., Zhai, G., Yang, X., and Zhang, W., "A new no-reference stereoscopic image quality assessment based on ocular dominance theory and degree of parallax," in [*Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*], 206–209, IEEE (2012).

[23] Wang, Z., Sheikh, H. R., and Bovik, A. C., "No-reference perceptual quality assessment of JPEG compressed images," in [*International Conference on Image Processing*], **1**, IEEE (2002).

[24] Akhter, R., Sazzad, Z. P., Horita, Y., and Baltes, J., "No-reference stereoscopic image quality assessment," in [*Stereoscopic Displays and Applications XXI*], **7524**, International Society for Optics and Photonics (2010).

[25] Ryu, S. and Sohn, K., "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Transactions on Circuits and Systems for Video Technology* **24**(4), 591–602 (2013).

[26] Ryu, S., Kim, D. H., and Sohn, K., "DIML Stereo video databases," (2012).

[27] Su, C.-C., Cormack, L. K., and Bovik, A. C., "Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation," *IEEE Transactions on Image Processing* **24**(5), 1685–1699 (2015).

[28] Shao, F., Li, K., Lin, W., Jiang, G., and Yu, M., "Using binocular feature combination for blind quality assessment of stereoscopic images," *IEEE Signal Processing Letters* **22**(10), 1548–1551 (2015).

[29] Shao, F., Lin, W., Wang, S., Jiang, G., Yu, M., and Dai, Q., "Learning receptive fields and quality lookups for blind quality assessment of stereoscopic images," *IEEE Transactions on Cybernetics* **46**(3), 730–743 (2015).

[30] Shao, F., Li, K., Lin, W., Jiang, G., and Dai, Q., "Learning blind quality evaluator for stereoscopic images using joint sparse representation," *IEEE Transactions on Multimedia* **18**(10), 2104–2114 (2016).

[31] Khan, Z. A., Kaaniche, M., Beghdadi, A., and Cheikh, F. A., "Joint statistical models for no-reference stereoscopic image quality assessment," in [*7th European Workshop on Visual Information Processing (EUVIP)*], 1–5, IEEE (2018).

[32] Wang, X. and Sheng, Y., "No-reference stereoscopic image quality assessment based on visual saliency region," in [*Chinese Automation Congress (CAC)*], 2070–2074, IEEE (2019).

[33] Appina, B., "A 'complete blind'no-reference stereoscopic image quality assessment algorithm," in [*International Conference on Signal Processing and Communications (SPCOM)*], 1–5, IEEE (2020).

[34] Chen, M.-J., Cormack, L. K., and Bovik, A. C., "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing* **22**(9), 3379–3391 (2013).

[35] Sun, G., Shi, B., Chen, X., Krylov, A. S., and Ding, Y., "Learning local quality-aware structures of salient regions for stereoscopic images via deep neural networks," *IEEE Transactions on Multimedia* **22**(11), 2938–2949 (2020).

[36] Shen, L., Fang, R., Yao, Y., Geng, X., and Wu, D., "No-reference stereoscopic image quality assessment based on image distortion and stereo perceptual information," *IEEE Transactions on Emerging Topics in Computational Intelligence* **3**(1), 59–72 (2018).

[37] Shao, F., Tian, W., Lin, W., Jiang, G., and Dai, Q., "Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions," *IEEE Transactions on Image Processing* **25**(5), 2059–2074 (2016).

[38] Appina, B., Khan, S., and Channappayya, S. S., "No-reference stereoscopic image quality assessment using natural scene statistics," *Signal Processing: Image Communication* **43**, 1–14 (2016).

[39] Oh, H., Ahn, S., Kim, J., and Lee, S., "Blind deep S3D image quality evaluation via local to global feature aggregation," *IEEE Transactions on Image Processing* **26**(10), 4923–4936 (2017).

[40] Yang, J., Sim, K., Lu, W., and Jiang, B., "Predicting stereoscopic image quality via stacked auto-encoders based on stereopsis formation," *IEEE Transactions on Multimedia* **21**(7), 1750–1761 (2018).

[41] Yang, J., Sim, K., Gao, X., Lu, W., Meng, Q., and Li, B., "A blind stereoscopic image quality evaluator with segmented stacked autoencoders considering the whole visual perception route," *IEEE Transactions on Image Processing* **28**(3), 1314–1328 (2018).

[42] Ding, Y., Deng, R., Xie, X., Xu, X., Zhao, Y., Chen, X., and Krylov, A. S., "No-reference stereoscopic image quality assessment using convolutional neural network for adaptive feature extraction," *IEEE Access* **6**, 37595–37603 (2018).

[43] Sheikh, H. R., Sabir, M. F., and Bovik, A. C., "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing* **15**(11), 3440–3451 (2006).

[44] Messai, O., Hachouf, F., and Seghir, Z. A., "Deep learning and cyclopean view for no-reference stereoscopic image quality assessment," in [*International Conference on Signal, Image, Vision and their Applications (SIVA)*], 1–6, IEEE (2018).

[45] Zhou, W., Chen, Z., and Li, W., "Dual-stream interactive networks for no-reference stereoscopic image quality assessment," *IEEE Transactions on Image Processing* **28**(8), 3946–3958 (2019).

[46] Kim, J., Ahn, S., Oh, H., and Lee, S., "CNN-Based Blind Quality Prediction On Stereoscopic Images Via Patch To Image Feature Pooling," in [*International Conference on Image Processing (ICIP)*], 1745–1749, IEEE (2019).

[47] Xu, X., Shi, B., Gu, Z., Deng, R., Chen, X., Krylov, A. S., and Ding, Y., "3D no-reference image quality assessment via transfer learning and saliency-guided feature consolidation," *IEEE Access* **7**, 85286–85297 (2019).

[48] Zhao, P., Li, S., and Chang, Y., "No-reference stereoscopic image quality assessment based on dilation convolution," in [*Visual Communications and Image Processing (VCIP)*], 1–4, IEEE (2019).

[49] Liu, T.-J., Lin, C.-T., Liu, H.-H., and Pei, S.-C., "Blind stereoscopic image quality assessment based on hierarchical learning," *IEEE Access* **7**, 8058–8069 (2019).

[50] Li, S. and Wang, M., "No-Reference Stereoscopic Image Quality Assessment Based on Convolutional Neural Network with A Long-Term Feature Fusion," in [*International Conference on Visual Communications and Image Processing (VCIP)*], 318–321, IEEE (2020).

[51] Sim, K., Yang, J., Lu, W., and Gao, X., "Blind stereoscopic image quality evaluator based on binocular semantic and quality channels," *IEEE Transactions on Multimedia* (2021).

[52] Galkandage, C., Calic, J., Dogan, S., and Guillemaut, J.-Y., "Stereoscopic video quality assessment using binocular energy," *IEEE Journal of Selected Topics in Signal Processing* **11**(1), 102–112 (2016).

[53] Appina, B. and Channappayya, S. S., "Full-reference 3-D video quality assessment using scene component statistical dependencies," *IEEE Signal Processing Letters* **25**(6), 823–827 (2018).

[54] Urvoy, M., Barkowsky, M., Cousseau, R., Koudota, Y., Ricorde, V., Le Callet, P., Gutierrez, J., and Garcia, N., "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," in [*Fourth International Workshop on Quality of Multimedia Experience*], 109–114, IEEE (2012).

[55] Appina, B., Manasa, K., and Channappayya, S. S., "Subjective and objective study of the relation between 3D and 2D views based on depth and bitrate," *Electronic Imaging* **2017**(5), 145–150 (2017).

[56] Md, S. K. and Channappayya, S., "Full Reference Stereoscopic Video Quality Assessment Based On Spatio-Depth Saliency And Motion Strength," in [*National Conference on Communications (NCC)*], 1–5, IEEE (2019).

[57] Galkandage, C., Calic, J., Dogan, S., and Guillemaut, J.-Y., "Full-reference stereoscopic video quality assessment using a motion sensitive hvs model," *IEEE Transactions on Circuits and Systems for Video Technology* (2020).

[58] De Silva, V., Arachchi, H. K., Ekmekcioglu, E., and Kondoz, A., "Toward an impairment metric for stereoscopic video: A full-reference video quality metric to assess compressed stereoscopic video," *IEEE Transactions on Image Processing* **22**(9), 3392–3404 (2013).

[59] Wang, J., Wang, S., and Wang, Z., "Asymmetrically compressed stereoscopic 3D videos: Quality assessment and rate-distortion performance evaluation," *IEEE Transactions on Image Processing* **26**(3), 1330–1343 (2017).

[60] Appina, B., Manasa, K., and Channappayya, S. S., "A full reference stereoscopic video quality assessment metric," in [*International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 2012–2016, IEEE (2017).

[61] Hong, W. and Yu, L., "A spatio-temporal perceptual quality index measuring compression distortions of three-dimensional video," *IEEE Signal Processing Letters* **25**(2), 214–218 (2017).

[62] Regis, C. D. M., de Miranda Cardoso, J. V., de Pontes Oliveira, Í., and de Alencar, M. S., "Objective estimation of 3D video quality: A disparity-based weighting strategy," in [*International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*], 1–6, IEEE (2013).

[63] Qi, F., Zhao, D., Fan, X., and Jiang, T., "Stereoscopic video quality assessment based on visual attention and just-noticeable difference models," *Signal, Image and Video Processing* **10**(4), 737–744 (2016).

[64] Malekmohamadi, H., Fernando, W., and Kondoz, A. M., "A new reduced reference objective quality metric for stereoscopic video," in [*Globecom Workshops*], 1325–1328, IEEE (2012).

[65] Appuhami, H. D., Martini, M. G., and Hewage, C. T., "Using 3D structural tensors in quality evaluation of stereoscopic video," in [*Visual Communications and Image Processing Conference*], 418–421, IEEE (2014).

[66] Hewage, C. T. and Martini, M. G., "Quality of experience for 3D video streaming," *IEEE Communications Magazine* **51**(5), 101–107 (2013).

[67] Hewage, C. T., Martini, M. G., Brandas, M., and De Silva, D. V. S., "A study on the perceived quality of 3D video subject to packet losses," in [*International Conference on Communications Workshops (ICC)*], 662–666, IEEE (2013).

[68] Yu, M., Zheng, K., Jiang, G., Shao, F., and Peng, Z., "Binocular perception based reduced-reference stereo video quality assessment method," *Journal of Visual Communication and Image Representation* **38**, 246–255 (2016).

[69] Sazzad, Z. P., Yamanaka, S., and Horita, Y., "Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction," in [*Second International Workshop on Quality of Multimedia Experience (QoMEX)*], 106–111, IEEE (2010).

[70] Solh, M. and AlRegib, G., "A no-reference quality measure for DIBR-based 3D videos," in [*International Conference on Multimedia and Expo*], 1–6, IEEE (2011).

[71] Ha, K. and Kim, M., "A perceptual quality assessment metric using temporal complexity and disparity information for stereoscopic video," in [*18th International Conference on Image Processing*], 2525–2528, IEEE (2011).

[72] Han, Y., Yuan, Z., and Muntean, G.-M., "No reference objective quality metric for stereoscopic 3D video," in [*International Symposium on Broadband Multimedia Systems and Broadcasting*], 1–6, IEEE (2014).

[73] ITU, T., "Opinion model for video-telephony applications," *ITU-T Recommendation P. 1070* (2007).

[74] Han, Y., Yuan, Z., and Muntean, G.-M., "Extended no reference objective quality metric for stereoscopic 3D video," in [*International Conference on Communication Workshop (ICCW)*], 1729–1734, IEEE (2015).

[75] "no-reference quality assessment of 3d videos based on human visual perception,"

[76] Cheng, E., Burton, P., Burton, J., Joseski, A., and Burnett, I., "RMIT3DV: Pre-announcement of a creative commons uncompressed HD 3D video database," in [*Fourth International Workshop on Quality of Multimedia Experience*], 212–217, IEEE (2012).

[77] Goldmann, L., De Simone, F., and Ebrahimi, T., "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," in [*Three-Dimensional Image Processing (3DIP) and Applications*], **7526**, International Society for Optics and Photonics (2010).

[78] Mahmood, S. A. and Ghani, R. F., "Objective quality assessment of 3D stereoscopic video based on motion vectors and depth map features," in [*7th Computer Science and Electronic Engineering Conference (CEEC)*], 179–183, IEEE (2015).

[79] Goldmann, L., De Simone, F., and Ebrahimi, T., "Impact of acquisition distortion on the quality of stereoscopic images," in [*Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics*], (CONF) (2010).

[80] Qi, F., Jiang, T., Fan, X., Ma, S., and Zhao, D., "Stereoscopic video quality assessment based on stereo just-noticeable difference model," in [*International Conference on Image Processing*], 34–38, IEEE (2013).

[81] Mittal, A., Soundararajan, R., and Bovik, A. C., "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters* **20**(3), 209–212 (2012).

[82] Appina, B., Dendi, S. V. R., Manasa, K., Channappayya, S. S., and Bovik, A. C., "Study of subjective quality and objective blind quality prediction of stereoscopic videos," *IEEE Transactions on Image Processing* **28**(10), 5027–5040 (2019).

[83] Yang, J., Zhao, Y., Jiang, B., Meng, Q., Lu, W., and Gao, X., "No-reference quality assessment of stereoscopic videos with inter-frame cross on a content-rich database," *IEEE Transactions on Circuits and Systems for Video Technology* **30**(10), 3608–3623 (2019).

[84] Yang, J., Zhao, Y., Jiang, B., Lu, W., and Gao, X., "No-reference quality evaluation of stereoscopic video based on spatio-temporal texture," *IEEE Transactions on Multimedia* **22**(10), 2635–2644 (2019).

[85] Hou, Y., Liu, L., Zhang, Y., and Sang, Q., "Stereoscopic video quality assessment using oriented local gravitational force statistics," *IEEE Access* **8**, 212442–212455 (2020).

[86] Xu, M., Chen, J., Wang, H., Liu, S., Li, G., and Bai, Z., "C3dvqa: Full-reference video quality assessment with 3d convolutional neural network," in [*ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 4447–4451, IEEE (2020).

[87] Seshadrinathan, K., Soundararajan, R., Bovik, A. C., and Cormack, L. K., "Study of subjective and objective quality assessment of video," *IEEE transactions on Image Processing* **19**(6), 1427–1441 (2010).

[88] Vu, P. V. and Chandler, D. M., "ViS3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices," *Journal of Electronic Imaging* **23**(1) (2014).

[89] Dendi, S. V. R., Krishnappa, G., and Channappayya, S. S., "Full-reference video quality assessment using deep 3d convolutional neural networks," in [*National Conference on Communications (NCC)*], 1–5, IEEE (2019).

[90] De Simone, F., Tagliasacchi, M., Naccari, M., Tubaro, S., and Ebrahimi, T., "A H. 264/AVC video database for the evaluation of quality metrics," in [*International Conference on Acoustics, Speech and Signal Processing*], 2430–2433, IEEE (2010).

[91] Moorthy, A. K., Choi, L. K., Bovik, A. C., and De Veciana, G., "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE Journal of Selected Topics in Signal Processing* **6**(6), 652–671 (2012).

[92] Yang, F., Wan, S., Xie, Q., and Wu, H. R., "No-reference quality assessment for networked video via primary analysis of bit stream," *IEEE Transactions on Circuits and Systems for Video Technology* **20**(11), 1544–1554 (2010).

[93] Yang, J., Zhu, Y., Ma, C., Lu, W., and Meng, Q., "Stereoscopic video quality assessment based on 3D convolutional neural networks," *Neurocomputing* **309**, 83–93 (2018).

[94] Zhou, W., Chen, Z., and Li, W., "Stereoscopic video quality prediction based on end-to-end dual stream deep neural networks," in [*Pacific Rim Conference on Multimedia*], 482–492, Springer (2018).

[95] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems* **25**, 1097–1105 (2012).

[96] Varga, D. and Szirányi, T., "No-reference video quality assessment via pretrained CNN and LSTM networks," *Signal, Image and Video Processing* **13**(8), 1569–1576 (2019).

[97] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., "Rethinking the inception architecture for computer vision," in [*Conference on Computer Vision and Pattern Recognition*], 2818–2826 (2016).

[98] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A., "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," in [*Conference on Artificial Intelligence*], **31**(1) (2017).

[99] Hosu, V., Hahn, F., Jenadeleh, M., Lin, H., Men, H., Szirányi, T., Li, S., and Saupe, D., "The Konstanz natural video database (KoNViD-1k)," in [*Ninth International Conference on Quality of Multimedia Experience (QoMEX)*], 1–6, IEEE (2017).

[100] Varga, D., "No-reference video quality assessment based on the temporal pooling of deep features," *Neural Processing Letters* **50**(3), 2595–2608 (2019).

[101] Feng, Y., Li, S., and Chang, Y., "Multi-Scale Feature-Guided Stereoscopic Video Quality Assessment Based on 3D Convolutional Neural Network," in [*International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 2095–2099, IEEE (2021).

[102] Banitalebi-Dehkordi, A., Pourazad, M. T., and Nasiopoulos, P., "3D video quality metric for mobile applications," in [*International Conference on Acoustics, Speech and Signal Processing*], 3731–3735, IEEE (2013).

[103] Hupont, I., Gracia, J., Sanagustin, L., and Gracia, M. A., "How do new visual immersive systems influence gaming QoE? A use case of serious gaming with Oculus Rift," in [*Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*], 1–6, IEEE (2015).

[104] Costa, L. C., Correa, A. G., José, M. A., Lotto, E. P., Martinazzo, A. A., Biazon, L. C., Ficheman, I. K., Nagamura, M., Lopes, R. D., and Zuffo, M. K., "3D stereoscopic game development technique for smart TVs," in [*International Conference on Consumer Electronics (ICCE)*], 1–2, IEEE (2014).

[105] Mahoney, N., Oikonomou, A., and Wilson, D., "Stereoscopic 3D in video games: A review of current design practices and challenges," in [*16th International Conference on Computer Games (CGAMES)*], 148–155, IEEE (2011).

[106] Malyshev, K., Lavrushkin, S., and Vatolin, D., "Stereoscopic Dataset from A Video Game: Detecting Converged Axes and Perspective Distortions in S3D Videos," in [*International Conference on 3D Immersion (IC3D)*], 1–7, IEEE (2020).

[107] Benzeroual, K. and Allison, R. S., "Cyber (motion) sickness in active stereoscopic 3D gaming," in [*International Conference on 3D Imaging*], 1–7, IEEE (2013).

[108] Kobayashi, N., Yamazaki, H., Ishikawa, M., and Momose, Y., "Effects of visual induced motion sickness of stereoscopic 3D interactive video," in [*4th Global Conference on Consumer Electronics (GCCE)*], 664–665, IEEE (2015).

[109] Nam, K. W., Park, J., Kim, I. Y., and Kim, K. G., "Application of stereo-imaging technology to medical field," *Healthcare informatics research* **18**(3), 158 (2012).

[110] Rodriguez-Palacios, A., Kodani, T., Kaydo, L., Pietropaoli, D., Corridoni, D., Howell, S., Katz, J., Xin, W., Pizarro, T. T., and Cominelli, F., "Stereomicroscopic 3D-pattern profiling of murine and human intestinal inflammation reveals unique structural phenotypes," *Nature communications* **6**(1), 1–16 (2015).

[111] Deng, K., Wei, B., Chen, M., Huang, Z., and Wu, H., "Realization of real-time X-ray stereoscopic vision during interventional procedures," *Scientific reports* **8**(1), 1–10 (2018).

[112] Wellens, L. M., Meulstee, J., van de Ven, C. P., van Scheltinga, C. T., Littooij, A. S., van den Heuvel-Eibrink, M. M., Fiocco, M., Rios, A. C., Maal, T., and Wijnen, M. H., "Comparison of 3-dimensional and augmented reality kidney models with conventional imaging data in the preoperative assessment of children with wilms tumors," *JAMA network open* **2**(4) (2019).

[113] Li, M., Ren, Y., and Weng, G., "Clinical study of three-dimensional laparoscopic partial nephrectomy for the treatment of highly complex renal tumors with renal nephrometry scores of ≥10 points," *BioMed Research International* **2020** (2020).

[114] Deepika, N. and Variyar, V. S., "Obstacle classification and detection for vision based navigation for autonomous driving," in [*International Conference on Advances in Computing, Communications and Informatics (ICACCI)*], 2092–2097, IEEE (2017).

[115] Wang, Y., Chao, W.-L., Garg, D., Hariharan, B., Campbell, M., and Weinberger, K. Q., "Pseudo-lidar from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving," in [*Conference on Computer Vision and Pattern Recognition*], 8445–8453 (2019).

[116] Wah, C., "Parking space vacancy monitoring," *Projects in Vision and Learning* (2009).