IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 16, 2018, accepted December 5, 2018, date of publication December 11, 2018,
date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2885818

2D and 3D Image Quality Assessment: A Survey of Metrics and Challenges

YUZHEN NIU[1,2,3], (Member, IEEE), YINI ZHONG[1], WENZHONG GUO[1,2], (Member, IEEE), YIQING SHI[1], AND PEIKUN CHEN[1]

[1]College of Mathematics and Computer Science, Fuzhou University, Fujian 350116, China
[2]Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou University, Fujian 350116, China
[3]Key Laboratory of Spatial Data Mining & Information Sharing, Ministry of Education, Fujian 350002, China

Corresponding author: Wenzhong Guo (fzugwz@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672158, Grant 61672159, Grant U1705262, and Grant 61502105, in part by the Technology Guidance Project of Fujian Province under Grant 2017H0015, in part by the Industry-Academy Cooperation Project of Fujian Province under Grant 2017H6008, in part by the Natural Science Foundation of Fujian Province under Grant 2018J01795, and in part by the Fujian Collaborative Innovation Center for Big Data Application in Governments.

ABSTRACT Image quality is important not only for the viewing experience, but also for the performance of image processing algorithms. Image quality assessment (IQA) has been a topic of intense research in the fields of image processing and computer vision. In this paper, we first analyze the factors that affect two-dimensional (2D) and three-dimensional (3D) image quality, and then provide an up-to-date overview on IQA for each main factor. The main factors that affect 2D image quality are fidelity and aesthetics. Another main factor that affects stereoscopic 3D image quality is visual comfort. We also describe the IQA databases and give the experimental results on representative IQA metrics. Finally, we discuss the challenges for IQA, including the influence of different factors on each other, the performance of IQA metrics in real applications, and the combination of quality assessment, restoration, and enhancement.

INDEX TERMS Image quality assessment, image aesthetics assessment, visual comfort, and image quality enhancement.

I. INTRODUCTION

The growing demand for better visual experience is driving the development and supply of multimedia shooting and display devices such as smart phones, digital cameras, and camcorders. The demand is also driving the continuous advance of technologies like multimedia and cloud storage, as well as the emergence and popularity of social media platforms like Facebook and WeChat. Images and videos have progressed from black and white to color; from analog signal to digital signal; from standard definition to high definition, super definition, blue-ray, 4K and 8K; and from two-dimensional (2D) to three-dimensional (3D), free viewpoint, virtual reality, and augmented reality.

The number of images and videos on the Internet is growing rapidly. However, not all images and videos on the Internet are guaranteed to be of high quality. At various stages of shooting, processing, storage, encoding, transmission, display, etc., the image and video quality is affected owing to distortion, destruction of composition, and uncomfortable depth, resulting in low quality images and videos. In addition, the differences between professional photographers and amateurs lead to large differences in the quality of captured images and videos.

Image quality is important not only for the viewing experience, but also for the performance of image processing algorithms. Various studies have found that distortion can affect the performance of algorithms such as face detection and recognition [1], image saliency detection [2], [3], event detection [4], and video target tracking [5]. Therefore, image quality assessment (IQA) has been a topic of intense research in the fields of image processing and computer vision.

Researchers have proposed various IQA metrics, many of which have found important applications in image processing systems. For example, IQA metrics can help image retrieval systems filter low-quality images by monitoring image quality, resulting in better subjective experience. Second, IQA metrics can determine whether the quality of an image will affect its subsequent usage, such as image quality in a data transmission network. Finally, IQA metrics can be used to measure the performance of quality restoration and enhancement algorithms. For example, IQA metrics can be introduced into image processing algorithms such as image super

(a) (b)

FIGURE 1. Images with different visual qualities. (a) A medium quality image. (b) A high quality image.

resolution [6], image deblurring [7], image denoising [8], and image deblocking [9]. The quality assessment result for each result image of the image processing algorithms can serve as a performance indicator.

The factors that affect the visual quality of 2D images are primarily low fidelity owning to distortion and poor aesthetics. Many types of distortion, such as noise, blur, compression, and packet loss, may be introduced in 2D images during various stages from shooting to display. However, not all undistorted 2D images have high visual quality (see Fig. 1(a)). The aesthetics of images also have a significant impact on whether or not 2D images are attractive. As shown in Fig. 1 (b), professional photographers usually follow the rules of composition, such as the rule of thirds and the rule of simplicity, to capture high-fidelity and attractive 2D images.

Stereoscopic 3D media provides users with 3D description of the real scenes using depth information from disparity between the left and right retinas, resulting in an immersive visual experience. 3D media has begun to expand from the film industry to a wider range of fields, including television entertainment, advertising, virtual reality, video surveillance, health care, military defense, games, design, manufacturing, and education. However, there are more factors that affect the visual quality of 3D media [10], [11] than those of 2D media, such as fidelity and aesthetics of left and right views, consistency between left and right views [12], visual comfort, screen quality and size, quality of 3D glasses, viewing distance and angle, physiological and psychological status of viewers, and so on.

The main factor that affects stereoscopic 3D image quality is visual comfort, besides fidelity and aesthetics. Distortion-free and highly appealing 3D images may still be considered to be of low-quality if they have low visual comforts. Fig. 2 shows an image whose pixels have non-positive disparity values, implying that all image contents are perceived to be in front of the display screen. The main objects in Fig. 2 are shown in front of the display screen with large absolute disparity values which are out of the comfort zone. Furthermore, the shoulder in Fig. 2 touches the left boundary in the right view, which leads to the window violation problem. Both, non-positive disparity values and window violation problem, usually make a 3D image uncomfortable for viewing and cause visual fatigue.

In this paper, we first provide an up-to-date overview on IQA for each main factor in Section II, including fidelity, aesthetics, and visual comfort. We then describe the IQA databases and the experimental results of representative IQA metrics in Section III. Finally, we discuss the challenges for IQA in Section IV, including the influence of different factors on each other, the performance of IQA metrics in real applications, and the combination of quality assessment, restoration, and enhancement. We conclude the paper in Section V.

II. 2D AND 3D IMAGE QUALITY ASSESSMENT METRICS

Existing IQA metrics assess the quality of 2D and 3D images from the perspectives of fidelity, aesthetics, and visual comfort. We give the up-to-date overviews on 2D and 3D fidelity assessment, aesthetics assessment, and visual comfort assessment (VCA) in Subsections II-A, II-B, II-C, and II-D, respectively.

A. 2D IMAGE FIDELITY ASSESSMENT

Distortions of a 2D image, such as noise, blur, and compression, usually lead to the reduction of image fidelity. Therefore, there are numerous works on 2D image fidelity assessment. Existing 2D image fidelity assessment metrics can be classified into three categories according to whether the reference image is required: full-reference (FR), reduced-reference (RR), and no-reference (NR). FR metrics require both distorted and reference images. RR metrics use only partial information of the reference image. For NR metrics, only the distorted image is necessary. FR metrics are mainly used to evaluate the performance of image processing algorithms. RR metrics are mainly used to access the distortion of image in the transmission process. NR metrics are mainly used to filter low-fidelity images in applications. We give the

(a) (b) (c)

FIGURE 2. A distortion-free 3D image that has a low subjective quality owing to its non-positive disparity map and window violation problem at the left image boundary. (a) Left view. (b) Right view. (c) Horizontal disparity map.

overviews on FR, RR, and NR 2D image fidelity assessment metrics in Subsections II-A1, II-A2 and II-A3, respectively.

1) FULL-REFERENCE 2D IMAGE FIDELITY ASSESSMENT

Conventional error quantification-based FR metrics, such as mean squared error (MSE), signal-to-noise ratio (SNR), and peak SNR (PSNR), do not consider the content of an image and the characteristics of human visual system (HVS). So these metrics usually result in a weak consistency with human perception.

Therefore, HVS-inspired and image content-aware fidelity assessment metrics have been presented. Researchers combined the characteristics of HVS with pure mathematical algorithms. Visual SNR (VSNR) [13] quantifies the visual fidelity of distorted images. HVS-based peak SNR (PSNR-HVS) [14] combines PSNR with HVS characteristics by considering the contrast sensitivity function (CSF). PSNRHA [15] introduces CSF, between-coefficient contrast masking of discrete cosine transform (DCT) basis functions, and mean shift into PSNR. PSNR-HMA [15] introduces contrast changing into PSNR-HA.

Wang and Bovik claimed that human eyes obtain image information through three channels, brightness, contrast, and structure [16], and developed a universal image quality index (UQI) [17] and a structural similarity (SSIM) [16] index. Thereafter, variants of SSIM-based metrics have been proposed, such as multi-scale SSIM (MS-SSIM) [18], modified SSIM with automatic down-sampling (MSSIM) [19], complex wavelets-SSIM (CW-SSIM) [20], information content weighted SSIM measure (IW-SSIM) [21], edge strength similarity-based index (ESSIM) [22], and the combination of optimal scale selection (OSS) model [23] and SSIM (OSS-SSIM).

Unlike most IQA metrics that adopt a single most relevant strategy used by HVS to assess image fidelity, the index presented by Larson and Chandler, named most apparent distortion (MAD) [24], is modeled using two separate strategies: a detection-based strategy and an appearance-based strategy. Zhang *et al.* [25] proposed the feature similarity index (FSIM), which uses phase congruency and gradient magnitude as the primary and secondary features. Gradient magnitude similarity deviation (GMSD) [26] was proposed based on the observation that image gradients are sensitive to image distortions. GMSD predicts image quality based on the gradient magnitude similarity between the reference and the distorted images. A visual saliency-induced index (VSI) [27] introduces visual saliency into assessment based on the observation that image saliency [28]–[31] plays an important role in image quality assessment. Visual saliency is first used to calculate the difference between the distorted and the reference images as a reference factor. During the final pooling, the saliency value at each pixel is used as a weighting value, which affects the importance of different pixels. DCT subbands similarity (DSS) [32] index uses changes in the structural information of the subbands in the DCT domain and the weighted quality estimates for these subbands to predict image quality. Nafchi *et al.* [33] proposed a mean deviation similarity index (MDSI) to assess image quality by considering gradient similarity and chromaticity similarity to measure structural and color distortions, respectively. The superpixel-based similarity index (SPSIM) [34] evaluates the overall visual impression on local images by calculating superpixel luminance similarity and superpixel chrominance similarity, then quantifies the structural variations by gradient similarity, and finally evaluates the image quality by combining the three types of features.

During the last few years, machine-learning methods have been applied to image fidelity assessment. These methods tackle image fidelity assessment in two steps: first design appropriate features and then perform regression or classification using the features. Narwaria and Lin proposed a singular value decomposition (SVD)-based metric using support vector regression (SVDR) [35]. SVDR uses SVD to extract singular vectors as features of reference and distorted images, and then uses support vector regression (SVR) for image fidelity prediction. Charrier *et al.* [36] proposed a metric named machine learning-based image quality measure (MLIQM). MLIQM first classifies a distorted image into five scales using multi-support vector machine (SVM) classification according to the quality scale recommended by the ITU [37], and then uses an SVR to predict the final score of the distorted image. Liu *et al.* [38] presented a multi-method fusion (MMF) metric which was motivated by the observation that no single methods can give the best performance in all distortion types. This metric first classifies the distortion type of distorted images into five categories and then combines multi image fidelity assessment metrics using SVR to predict the image quality scores. A machine learning-based metric with distortion measured by non-negative matrix factorization (NMF) was proposed in [39]. Metric NMF uses extreme learning machine (ELM) method to address the limitations of existing pooling techniques.

Because convolutional neural network (CNN) is capable of learning features and regression based on the raw image data, it has been introduced to solve image fidelity assessment problem. Bosse *et al.* [40] presented a weighted average deep image quality measure for FR-IQA (WaDIQaM-FR). It is purely data-driven and does not rely on hand-crafted features or other types of prior domain knowledge about HVS or image statistics. Kim and Lee [41] presented a deep image quality assessment (DeepQA) model which learns the visual sensitivity characteristics of the HVS by using a deep CNN.

2) REDUCED-REFERENCE 2D IMAGE FIDELITY ASSESSMENT

Ma *et al.* [42] presented an RR image fidelity assessment metric using reorganized DCT-based image representation. It extracts the statistical information of the DCT coefficients of natural images, and uses statistical methods to design image features for image fidelity prediction. Rehman and Zhou [43] proposed an RR metric using structural similarity estimation. They extracted statistical features from a multi-

scale multi-orientation divisive normalization transform, and developed a distortion measurement based on SSIM [16]. Wu *et al.* [44] presented an RR metric which first maps visual content into a histogram based on orientation-selectivity-based visual pattern, and then evaluates image fidelity by calculating the changes between the two histograms of reference and distorted images. The structure degradation based RR metric [45] relies on the reduction degree of spatial frequency response of different distortion types and quality levels. Gu *et al.* [46] proposed an RR metric for contrast change using phase congruency and statistics information of the image histogram.

3) NO-REFERENCE 2D IMAGE FIDELITY ASSESSMENT

In recent years, NR image fidelity assessment has become a research hotspot in image fidelity assessment, as it does not require pristine knowledge of reference images.

Some distortion-specific NR metrics have been presented to assess the fidelity of blurred images [47], JPEG compressed images [48], deblocked images [49], and contrast-changed images [50].

Many natural scene statistics (NSS)-based NR metrics [51]–[56] have been presented to assess the fidelity of images distorted by various distortion types. The distortion identification-based image verity and integrity evaluation (DIIVINE) [51] uses the Gaussian scale mixture (GSM) model to model a set of neighboring wavelet coefficients and performs a wavelet decomposition using a steerable pyramid over two scales and six orientations to extract NSS features. It then uses a 2-stage framework that first identifies the image distortion type and thereafter estimates the distortion-specific quality. The blind image integrity notator using DCT statistics (BLIINDS-II) [52] is a fast single-stage framework that relies on a statistical model of local DCT coefficients. DCT features are extracted from the NSS model and thereafter fed to the Bayesian probabilistic inference model to evaluate image fidelity. The blind/referenceless image spatial quality evaluator (BRISQUE) [53] extracts NSS features from a statistical model of locally normalized luminance coefficients in the spatial domain and demonstrates that these features correlate well with human judgments on quality.

Some machine-learning-based NR metrics have also been presented [57], [58]. The shearlet and stacked autoencoders-based NR image quality assessment (SESANIA) index [57] extracts features using a multi-scale directional transform (shearlet transform) and utilizes the sum of subband coefficient amplitudes as the primary features. The stacked autoencoders are then employed to obtain the evolved features. Finally, SESANIA formulates the NR image fidelity assessment problem as a classification problem and uses softmax classifier to determine image fidelity. The local gradient patterns-based index [58] first extracts local statistical features from the gradient magnitude and phase of the image, and then uses a SVR to map the subjective MOS to local statistical features that convey important structural information.

Recently, CNN has been successfully introduced to the field of image fidelity assessment. A pioneering work, IQA-CNN, was proposed by Kang *et al.* [59]. Subsequently, many efforts have been made to improve its performance by designing deeper convolutional structures. In particular, Bosse *et al.* [40] proposed a deep NR image quality measure (DIQaM-NR). It effectively improves the performance of NR image fidelity assessment compared with related shallow models. Kim *et al.* proposed a blind image evaluator based on a convolutional neural network (BIECON). BIECON has two steps: first, the image is divided into patches to train the CNN, and the features extracted from all patches of the same image are combined to generate the feature vector of the image. Then a regression is used to calculate the quality score of the image [60]. Pan *et al.* [61] proposed a blind predicting similar quality map (BPSQ) for image fidelity assessment, which uses an FR metric to calculate similar images of image patches (such as FSIM, SSIM, and MDSI), and then use an U-shaped full CNN to train a similar graph generation model. Then the model generates the similarity graph which is used to train a CNN regression network for predicting the quality score of the image patch. Lin and Wang [62] proposed a Hallucinated-IQA framework for NR image fidelity assessment via adversarial learning. Hallucinated-IQA jointly optimizes quality-aware generative network, hallucination-guided quality regression network, and IQA-discriminator in an end-to-end manner.

B. 3D IMAGE FIDELITY ASSESSMENT

Vision is one of the most important ways for people to get information from the outside world. Because humans see 3D scenes of nature, it has always being humans' pursuit to reproduce the real and natural 3D scenes on the screen. With the rapid development of 3D display technology, 3D media has been more and more widely used in peoples' daily life and work, such as 3D video, games, and virtual reality. Compared to a 2D image, a stereoscopic 3D image involves depth information and consists of two views. Therefore, 3D IQA is more complicated than 2D IQA. Because the researches on 3D image fidelity assessment mainly focus on FR and NN metrics, we give the overviews on FR and NR 3D image fidelity assessment metrics in Subsections II-B1 and II-B2, respectively.

1) FULL-REFERENCE 3D IMAGE FIDELITY ASSESSMENT

FR 3D image fidelity assessment compares the distorted 3D image with the reference 3D image. According to the type of information they utilizes, FR 3D image fidelity assessment metrics can be divided into three categories. The first category of FR 3D metrics only considers the two views of a 3D image. In [63] and [64], existing 2D image fidelity assessment metrics are applied to the left and right views of a 3D image and the computed fidelity scores are combined to obtain a final fidelity score. Because a 3D image is not only two 2D images, these straightforward 2D-based 3D metrics cannot achieve good performance.

The second category of FR 3D metrics regards depth/disparity information as an important factor related to stereoscopic visual perception and assesses both the quality of the 2D views and that of the depth information. Benoit *et al.* [65] proposed first using FR 2D metrics C4 [66] and SSIM [16] to assess the image fidelity and then combining the two 2D fidelity scores and depth information to produce a 3D fidelity score. You *et al.* [67] investigated several FR 2D metrics on 3D images and proposed a model that combines the SSIM scores of the left and right views and a quality score computed based on the original and the distorted disparity maps. The cyclopean multiscale SSIM (CMS-SSIM) index [68] applies a 2D MS-SSIM [18] index to the Gabor cyclopean images of the original and distorted images. The Gabor cyclopean image is computed as a weighted summation of the left view and corresponding right view which is shifted according to the disparity map. Zhan*et al.* [69] presented a method based on machine learning regression, which learns from the view quality features, the disparity quality features, and a 3D quality feature.

The third category of FR 3D metrics considers binocular perceptual properties. Bensalma and Larabi [70] proposed an index that estimates the image fidelity based on the binocular fusion. Shao *et al.* [71] presented an index from the perspective of learning binocular receptive field properties using sparse representation, which has demonstrated good performance in simulating cells in the primary visual cortex and characterizing their impacts on fidelity estimation. Wang *et al.* [72] first presented an information and distortion weighted SSIM (IDW-SSIM) index to estimate the fidelity of each single view. Then, a multi scale model inspired by binocular rivalry was proposed to predict the fidelity of a stereoscopic image based on the fidelity scores of the left and right views.

2) NO-REFERENCE 3D IMAGE FIDELITY ASSESSMENT

Because NR 3D image fidelity assessment metrics do not use any information from the original reference image, NR metrics have more practical application prospects than FR and RR metrics. Based on the information that NR 3D metrics used, NR 3D metrics can be further classified into three categories as follows: binocular perception-based metrics, depth perception-based metrics, and difference perception-based metrics.

Many binocular perception-based metrics have been proposed by incorporating binocular perception, such as binocular combination, binocular rivalry, and binocular suppression. Ryu and Sohn [73] proposed an index that measures the extents of blurriness and blockiness for the left and right views and then combines these extents using a binocular perception model. Shao *et al.* [74] developed a phase-tuned fidelity lookup and a visual codebook from the binocular energy responses to achieve blind fidelity prediction by pooling. Zhou *et al.* [75] first presented two binocular combinations of stimuli, generated by an eye-weighting model and a contrast-gain control model, and then used the ELM to perform a fidelity prediction. Shao *et al.* [76] presented a framework for NR 3D image fidelity assessment by combining feature-prior and feature-distribution, which characterizes feature-prior by SVR and implements feature-distribution by sparsity regularization as the basis of weights for binocular combination. Shao *et al.* [77] proposed a domain transfer framework for NR fidelity prediction of asymmetrically distorted 3D images. It transfers the information from the source feature domain to its target fidelity domain by dictionary learning.

Depth perception-based metrics assess image fidelity based on the disparity map or synthesized cyclopean (human brain) image. Akhter and Horita [78] proposed an NR index that first extracts image features from a 3D image and its disparity map, and then uses a logistic regression model to predict image fidelity. Chen *et al.* [79] proposed combining 2D cues [68] in the cyclopean view and 3D cues [67] in the disparity map to estimate the perceptual fidelity of 3D images. Jiang *et al.* [80] proposed an index based on deep non-negativity constrained sparse autoencoder using the cyclopean image, left view, and right view as the input.

Difference perception-based metrics assess image fidelity based on the difference between the left and right views. Zhang *et al.* [81] proposed a CNN-based NR index, which considers the difference image as the representation of the depth and distortion in a 3D image. The end-to-end CNN model performs both image feature extraction and image fidelity prediction. Shen *et al.* [82] proposed combining the spatial frequency information and statistic feature extracted from the cyclopean and difference map to represent the binocular characteristic and asymmetric information of a 3D image.

C. IMAGE AESTHETICS ASSESSMENT

A distortion-free image can only guarantee to have a high fidelity, but cannot guarantee to have a high aesthetic quality. When shooting an image, a professional photographer designs the composition, color, illumination, contrast, and other aesthetic aspects of the image according to the shooting content. In this way, the image captured by a professional photographer usually presents a strong sense of beauty and is attractive to the viewer. At present, many aesthetics assessment metrics for 2D images have been presented, which enable the computer to determine whether an image is beautiful or ugly.

Early presented image aesthetics assessment metrics first design features related to aesthetic quality, and then map features to aesthetic score using effective machine learning methods. Datta *et al.* [83] presented a 56-dimensional feature, which includes some low-level features such as color, texture, shape, and size of the image and some high-level features such as depth of field, the rule of thirds, and region contrast. Ke *et al.* [84], inspired by photographic knowledge, designed a 7-dimensional feature which is composed of high-level semantics features, including picture simplicity, clarity, color, contrast, average brightness (exposure), etc.

Luo and Tang [85] proposed first separating the foreground and background of the image, and then extracting a 5-dimensional feature based on the foreground-background contrast, including the sharpness contrast, brightness contrast, color simplicity, harmony, and conformity of the rule of thirds. Bhattacharya *et al.* [86] used the rule of thirds and the principle of visual balance to assess the aesthetic quality of the image and enhanced the aesthetic quality of the image according to these rules. Nishiyama *et al.* [87] proposed detecting a series of subject regions in the image and training an SVM classifier using the subject regions. Finally, the SVM scores of the subject regions are combined to calculate the aesthetic quality of the image. Nishiyama *et al.* [88] proposed extracting local color descriptors and constructing histograms as features for aesthetic image classification. Works in [89]–[91] directly use Fisher vector as feature to predict the aesthetic quality of an image. Su *et al.* [92] proposed directly using the bag of visual words as feature. Luo*et al.* [93] proposed a content-based image aesthetics assessment index, which divided the image into 7 categories and designed different features for different categories. Marchesotti *et al.* [94] proposed using visual attributes as feature to solve the problem of visual attractiveness analysis.

As designing hand-crafted feature is challenging, researchers also proposed some end-to-end aesthetics assessment metrics based on deep learning [95]–[97]. Because the CNN with full connection layers requires the input image of the network to have a fixed size, works [96], [97] first transform the input image using cropping, scaling, padding, etc. In order to avoid the influence of image transformation on aesthetics assessment. Mai *et al.* [98] proposed a multi-net adaptive spatial pooling (ASP) ConvNet architecture, which can directly process images of different sizes and aspect ratios. Kong *et al.* [99] proposed using AlexNet-inspired architecture [100] to predict different image attributes. They used a regression network to predict aesthetics rating. Then, they used the Siamese Network architecture [101] to jointly optimize the network to predict the aesthetic score and relative ranking of two images. Kucer *et al.* [102] selected the specific combinations of 25 and 40 hand-crafted features from more than 300 hand-crafted features and obtained good classification/regression performance. The method also integrates the selected hand-crafted features with the pre-trained deep learning features, which significantly improves the accuracy of classification and prediction of aesthetic score. Liu *et al.* [103] proposed a semi-supervised deep active learning (SDAL)-based aesthetics assessment metric. Different from [97] which randomly selects object blocks, this method successively connects the semantically important object blocks from each scene to learn the human gaze transfer path (GSP), thus better simulates human visual perception. Then a probabilistic model is developed for the image aesthetics assessment using the GSP features that are learned using deep learning.

The research on 3D image aesthetics assessment is very limited so far, which needs further investigation.

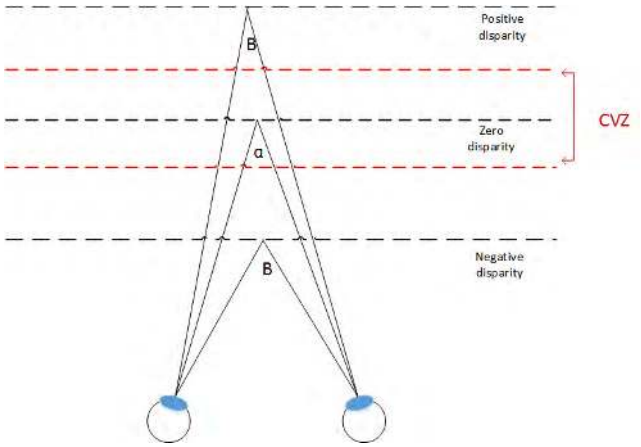

FIGURE 3. Geometry of stereoscopic display system. (This figure comes from [107]).

Niu *et al.* [104] proposed a 3D image resizing method which adapt a 3D image to different sizes and aspect ratios while retaining its aesthetics. This method measures the quality of resized images from an aesthetics point of view so as to best retain the aesthetic quality of the original 3D images in the resized results. Islam *et al.* [105] proposed a 3D deformation method based on aesthetics. This method reconstructs 3D images based on two common photographic composition rules, the rule of thirds and the rule of visual balance, but this method may destroy the shape of foreground objects in the resulting 3D images. In a follow-up work [106], Islam *et al.* proposed an aesthetics-driven 3D image recomposition method, which can adapt to the depth changes of foreground objects in a given 3D image to enhance human visual aesthetic experience.

D. 3D IMAGE VISUAL COMFORT ASSESSMENT

The most distinctive difference between 2D and 3D images is depth perception. The differences of the corresponding pixels in the left and right retinas in the horizontal and vertical directions are called the horizontal and vertical disparities. These two kinds of disparities are the main factors for visual comfort. Especially, the vertical disparity should be close to zero for a high quality stereoscopic 3D image. The horizontal disparity determines the depth perception of the 3D content. Excessive horizontal disparity will cause vergence-accommodation conflict [10], [11], therefore leads to visual fatigue for human eyes. Hereafter we use disparity instead of horizontal disparity when without introducing ambiguity. Fig. 3 shows the situations when objects are perceived in the negative disparity zone, positive disparity zone, and the comfort viewing zone (CVZ) which is close to the zero disparity through human's viewing system.

Comparing with the quality assessment metrics for image fidelity, there have less works focus on visual comfort assessment. Researchers first analyze the influences of depth on visual comfort using a series of subjective user studies. The experiments in [108] were conducted to analyze the effects of camera parameters of a 3D image and the duration of display on image perception, including the naturalness of image and

the image quality. The results show that during the shooting and processing of a 3D image, the toe-in of the two lenses should be avoided as much as possible. Work [109] evaluates the 3D image from six aspects subjectively, including the distortions, depth information, stereo visual comfort, whether the stereo feeling exits, naturalness of the stereo feeling, and whether suitable for viewing. Work [110] investigates the influence of dynamic fast disparity adjustment on visual comfort through the user subjective questionnaires. Visual comfort is reflected by the stereo fusion time of the image after disparity adjusted. This work found out that, when the disparity adjust range is large, controlling the users' focus fields to be close to the zero disparity can effectively reduce the stereo fusion time.

Based on the subjective experiments, researchers began to study objective depth VCA metrics [107], [111], [112]. Work [107] first calculates the disparity values of the matched feature points in the left and right views. It then computes the range of disparity, the average negative and positive disparity values, the average absolute disparity value, and then combines these features to train a visual comfort predictor using SVR. Work [111] takes the neural activity into account. It first extracts the disparity as the rough features to estimate the corresponding neural cell's activities which are called fine features. The final visual comfort predictor is trained using SVR based on both rough and fine features. These works [107], [111] need the image's mean opinion score (MOS) as the ground truth image visual comfort score. Because it takes a lot of time to obtain MOS, Jiang *et al.* [113] proposed an MOS-aware model for visual comfort predictor. It is based on the fact that the observers are much easier to make a preference judgment between two images. To facilitate further researches on VCA, Jiang *et al.* created a 3D image database NUB 3D-VCA. Inspired by the traditional absolute categorical rating (ACR) methodology in subjective study, work [114] proposed assessing visual comfort from the perspective of learning to rank. It first extracts some visual comfort-aware features and disparity features to train a comfort-aware function. The main goal is to find the optimal weighting vector that satisfies with the constraint between the ranks of 3D images and the comfort-aware function.

Inspired by the mechanism of HVS, specifically viewers usually pay more amount of attention to salient regions in an image, works [115]–[119] proposed some VCA metrics based on image saliency map. Work [115] proposed a visual comfort predictor using the 3D visual saliency. Firstly, it obtains the foreground region of the image and gets 3D visual saliency map based on both the disparity map and the saliency map for one view. Secondly, it calculates the weight values from the 3D visual saliency map and uses the weight values in the final predictor. According to the depth distribution of 3D images, work [116] designed an eye-tracking experiment to rank the disparity into different levels. Because humans pay different amounts of attention to different disparity levels, this work uses the Kullback-Leibler divergence to different disparity levels to obtain the corresponding saliency entropy and then adds it into the final visual comfort predictor. Work [117] also exploits human visual attention model for VCA. The overall framework consists of four steps. First, it computes the image-based and the depth-based saliency maps. Second, it uses the linear combination function on two saliency maps to obtain the final visual importance map. Then, it uses the visual importance map and the disparity map to extract a disparity feature vector. Finally, an overall comfort score prediction is trained using SVR based on the disparity feature vector. Because an image may contain multiple salient objects, work [118] created a multiple salient objects 3D image database (MSID) which contains more than two salient objects in each image. Then it extracts four kinds of features from the multiple salient objects, including disparity distribution feature, disparity jump feature, object distribution feature, as well as object width feature. Finally, it uses these features to train a visual comfort assessment model on MSID. Jiang *et al.* [119] utilized the fact that different neural activities effected by different depth levels. Owing to the fact that the neural activities of the middle temporal area of human brain are most sensitive to the stimulate response to human eyes, this work focuses on this situation and calculates the visual weights based on the disparity map. Based on the extracted features and their different weights, random forest regression algorithm is used to train a visual comfort predictor.

With the development of deep learning, some researchers used neural networks to assess visual comfort for 3D images. Work [112] presented a VCA index based on deep learning. The method firstly uses a twin neural network to extract the visual difference between the left and right views, and then use deep CNN to extract the disparity features based on HVS. Finally it utilizes the linear combination on these features and the human subjective scores to obtain a visual comfort predictor. Work [120] presented a novel binocular fusion deep network (BFN) for the prediction of visual discomfort, which takes account of the binocular fusion of the left and right views. And then the BFN is devised to learn the latent binocular features by fusing the spatial features of two views. They also devised a disparity regularization network (DRN) to consider the disparity characteristics. The proposed BFN and DRN are trained in a unified framework for both visual comfort prediction and disparity estimation. The final visual comfort score is predicted by the trained BFN without DRN.

III. EXPERIMENTS

In this section, we give the databases and experimental results on 2D and 3D image fidelity assessment, aesthetics assessment, and visual comfort assessment in Subsections III-A, III-B, III-C, and III-D, respectively.

A. EXPERIMENTAL RESULTS ON 2D IMAGE FIDELITY ASSESSMENT

We first describe the commonly used 2D image fidelity assessment databases. The information on these databases is summarized in Table 1.

FIGURE 4. Sample images in the LIVE database. (a) Reference. (b) WN. (c) JPEG. (d) JP2K. (e) GBLUR. (f) FF.

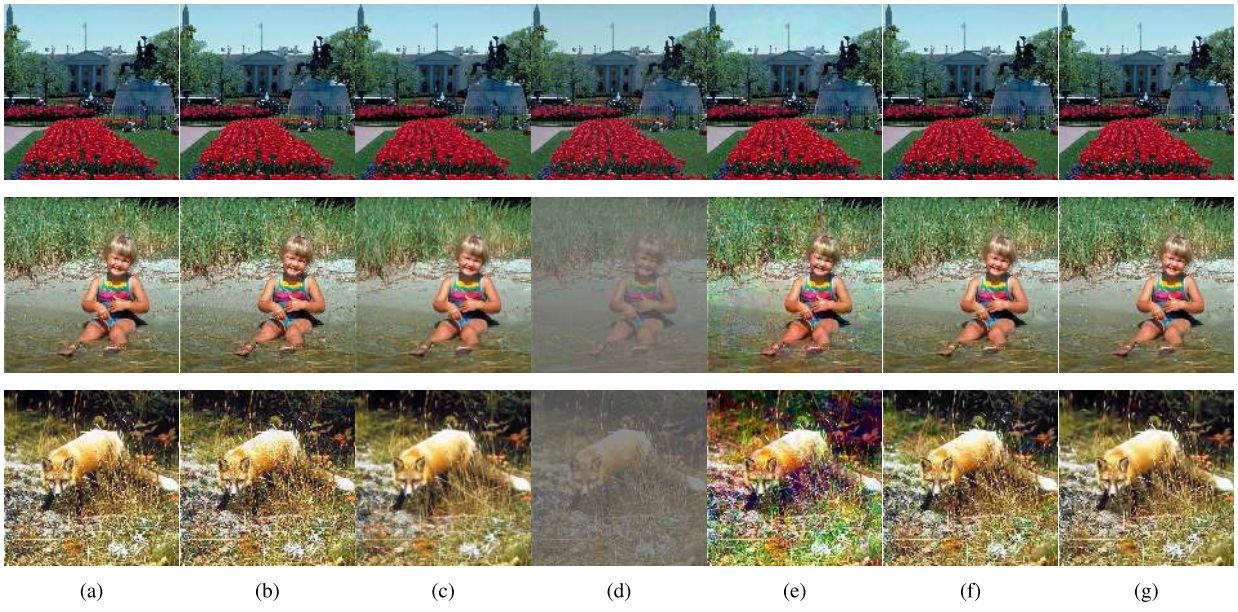

FIGURE 5. Sample images in the CSIQ database. (a) Reference. (b) AWGN. (c) GBLUR. (d) Contrast. (e) Fnoise. (f) JPEG. (g) JP2K.

TABLE 1. Summarized information on 2D image fidelity assessment databases.

Database	# of ref. images	# of dist. images	# of dist. types	Year	Ref.
LIVE	29	779	5	2005	[121]
CSIQ	30	866	6	2010	[24]
TID2008	25	1700	17	2009	[122]
TID2013	25	3000	24	2013	[123]

1) LIVE DATABASE [121]

This database contains 779 distorted images distorted by 5 distortion types: White Noise (WN), Gaussian blur (GBLUR), JPEG compression (JPEG), JPEG2000 compression (JP2K), and Fast-Fading (FF), at different distortion levels derived from 29 reference images. A differential mean opinion score (DMOS) is assigned to each distorted image, approximately in the range from 0 to 100. A higher DMOS indicates a lower fidelity of the image. Fig. 4 gives three examples for each distortion type. The distortion level increases for images from top to bottom for columns (b)-(f).

2) CSIQ DATABASE [24]

This database contains 866 distorted images distorted by 6 distortion types at 4 or 5 distortion levels derived from 30 reference images. The distortion types include additive white Gaussian noise (AWGN), GBLUR, global contrast decrements, additive pink Gaussian noise (fnoise), JPEG, and JP2K. The database contains 5000 subjective ratings reported in the form of DMOS in the range [0, 1] from 16 non-expert observers. Fig. 5 shows three examples for each distortion type. The distortion level increases for images from top to bottom for columns (b)-(g).

FIGURE 6. Sample images in the TID2013 database.

3) TID2008 DATABASE [122]

This database contains 1700 distorted images distorted by 17 distortion types at 4 distortion levels derived from 25 reference images. Each image is associated with an MOS in the range from 0 to 9. A higher MOS value indicates a higher fidelity of the image.

4) TID2013 DATABASE [123]

This database contains 3000 distorted images distorted by 24 distortion types at 5 distortion levels derived from 25 reference images. Each image is associated with an MOS in the range from 0 to 9. The first 17 distortion types in TID2013 database are shared with TID2008 database. Fig. 6 gives a reference image and its 24 distorted images, and one for each distortion type.

The performance indicators that are commonly adopted by image fidelity assessment metrics are Spearmans rank order correlation coefficient (SRCC), Pearsons linear correlation coefficient (PLCC), and root-mean-square error (RMSE). SRCC is used to measure the prediction monotonicity of an objective assessment index. PLCC is used to measure the relevance between the subjective evaluation and objective evaluation after nonlinear regression [124]. Larger PLCC and SRCC values indicate a closer relation with the human subjective evaluation. RMSE is used to assess the accuracy of the predictions after the nonlinear regression [124]. A smaller RMSE values indicate a superior correlation with human perception.

In Table 2, we show the experimental results of 21 FR, 5 RR, and 10 NR 2D image fidelity assessment metrics on the LIVE, CSIQ, TID2008, and TID2013 databases. As shown in Table 2, these metrics usually perform better on the LIVE and CSIQ databases than on the TID2008 and TID2013 databases because there are more distortion types in the TID2008 and TID2013 databases.

Early content-aware and HVS-based FR metrics [24], [26], [27], [32]–[34] have achieved good results on the LIVE and CSIQ databases. After that, the multi-strategy-based FR metrics [38], [39] have further improved the performance, especially on the TID2008 database. In recent years, the deep learning-based metrics [40], [41] have boosted the performance of the FR metrics on the TID2013 database. For RR

TABLE 2. Experimental results of existing 2D image fidelity assessment metrics on LIVE, CSIQ, TID2008, and TID2013 databases. The top three performance values for FR, RR, and NR metrics are formatted in bold and highlighted in red, green, and blue colors, respectively. And the symbol "-" indicates that the value is not provided in the corresponding paper and we could not find the corresponding source code.

Type	Metric	LIVE			CSIQ			TID2008			TID2013		
		SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE
FR	VSNR [13]	0.927	0.923	10.506	0.811	0.800	0.158	0.705	0.682	0.982	0.681	0.740	0.839
	PSNRHVS [14]	0.919	0.903	12.540	0.830	0.804	0.156	0.594	0.608	1.065	0.654	0.430	0.704
	PSNRHMA [15]	0.934	0.918	10.857	0.912	0.891	0.119	0.845	0.832	0.745	0.813	0.823	0.704
	SSIM [16]	0.948	0.845	8.946	0.876	0.861	0.133	0.775	0.773	0.851	0.742	0.790	0.761
	UQI [17]	0.894	0.899	11.982	0.810	0.831	0.146	0.585	0.664	1.003	-	-	-
	MS-SSIM [18]	0.951	0.949	8.169	0.913	0.899	0.115	0.854	0.845	0.717	0.786	0.833	0.686
	IW-SSIM [21]	0.957	0.952	8.347	0.921	0.914	0.106	0.856	0.858	0.690	0.778	0.832	0.688
	ESSIM [22]	0.962	0.671	15.828	0.933	0.642	0.167	0.884	0.841	0.777	0.804	0.828	0.694
	OSS-SSIM [23]	0.933	0.752	17.509	0.886	0.822	0.147	0.772	0.758	0.874	0.745	0.773	0.787
	MAD [24]	0.967	0.968	**6.907**	0.947	0.950	0.082	0.834	0.831	0.747	0.781	0.827	0.698
	FSIM [25]	0.963	0.960	7.678	0.924	0.912	0.108	0.881	0.874	0.653	0.802	0.859	0.635
	GMSD [26]	0.959	0.938	9.442	0.960	0.947	0.084	0.889	0.875	0.758	0.801	0.854	0.645
	VSI [27]	0.952	0.948	8.682	0.942	0.928	0.098	0.898	0.876	0.647	0.897	0.900	**0.540**
	DSS [32]	0.962	0.931	9.961	**0.961**	0.957	**0.076**	0.873	0.877	0.644	0.792	0.848	0.658
	MDSI [33]	0.967	0.966	7.079	0.957	0.953	0.080	**0.921**	**0.916**	**0.5314**	0.890	**0.909**	**0.5114**
	SPSIM [34]	0.962	0.960	7.629	0.944	0.934	0.093	0.910	0.893	0.605	**0.904**	**0.909**	**0.517**
	SVDR [35]	0.879	0.879	7.686	0.862	0.888	0.121	0.777	0.789	0.825	-	-	-
	CD-MMF [38]	**0.981**	**0.980**	**5.413**	**0.967**	**0.9614**	**0.067**	**0.942**	**0.9414**	**0.429**	-	-	-
	NMF [39]	**0.976**	**0.976**	**5.866**	**0.973**	**0.976**	**0.057**	**0.947**	**0.951**	**0.419**	-	-	-
	WaDIQaM-FR [40]	**0.970**	**0.980**	-	-	-	-	-	-	-	**0.940**	**0.946**	-
	DeepQA [41]	**0.981**	**0.982**	-	**0.961**	**0.965**	-	**0.947**	**0.951**	-	**0.939**	**0.947**	-
RR	RR-DCT [42]	**0.8814**	-	**12.830**	-	-	-	-	-	-	-	-	-
	RR-SSIM [43]	**0.913**	**0.919**	**9.189**	**0.853**	**0.843**	**0.141**	**0.721**	**0.723**	**0.927**	-	-	-
	OSVP [44]	-	-	-	**0.849**	**0.843**	**0.141**	-	-	-	**0.654**	**0.724**	**0.856**
	SDM [45]	**0.936**	**0.933**	-	-	-	-	-	-	-	-	-	-
	RIQMC [46]	-	-	-	**0.9514**	**0.965**	**0.044**	**0.810**	**0.859**	**0.493**	**0.804**	**0.865**	**0.492**
NR	DIIVINE [51]	0.916	0.917	-	-	-	-	-	-	-	-	-	-
	BLIINDS-II [52]	0.920	0.923	-	-	-	-	-	-	-	-	-	-
	SESANIA [57]	0.934	0.948	-	-	-	-	-	-	-	-	-	-
	LGP [58]	0.957	0.954	**7.901**	-	-	-	-	-	-	-	-	-
	BRISQUE [53]	0.940	0.942	-	-	-	-	-	-	-	-	-	-
	IQA-CNN [59]	0.956	0.953	-	-	-	-	-	-	-	-	-	-
	DIQaM-NR [40]	0.960	**0.972**	-	-	-	-	-	-	-	**0.835**	**0.855**	
	BIECON [60]	**0.961**	0.962	-	**0.825**	**0.8314**	-	-	-	-	0.721	**0.761**	-
	BPSQM [61]	**0.973**	**0.963**	-	**0.862**	**0.885**	-	-	-	-	**0.862**	**0.885**	-
	Hallucinated-IQA [62]	**0.982**	**0.982**	-	-	-	-	-	-	-	**0.879**	-	-

metrics, the performance on the LIVE and CSIQ databases is good. However the performance on the TID2008 and TID2013 databases needs further improvements. Before deep learning was introduced to NR metrics, previous NR metrics are much inferior to FR metrics. In recent years, the performance of NR metrics [40], [59]–[62] achieved large improvements.

B. EXPERIMENTAL RESULTS ON 3D IMAGE FIDELITY ASSESSMENT

We first describe the commonly used 3D image fidelity assessment databases. The information on these databases is summarized in Table 3.

1) LIVE 3D IQA DATABASE PHASE-I [125]

It contains 20 reference 3D images and 365 distorted 3D images, including 80 images for JPEG, JP2K, FF, and WN, and 45 images for GBLUR. Each image in the database is symmetrically distorted on its left and right views. Each distorted 3D image is given a DMOS from subjective evaluation. The values of DMOS are distributed in the interval of [-10, 100]. Fig. 7 shows a reference 3D image and its distorted images distorted by 5 distortion types.

TABLE 3. Summarized information on 3D image fidelity assessment databases.

Database	# of ref. images	# of dist. images	# of dist. types	Year	Ref.
LIVE 3D I	20	365	5	2013	[125]
LIVE 3D II	8	360	5	2013	[68]
NBU 3D	12	312	5	2013	[126]

2) LIVE 3D IQA DATABASE PHASE-II [68]

It contains 8 reference 3D images and 360 symmetrically or asymmetrically distorted 3D images. An asymmetrically distorted image has different distortion levels in its left and right views. The distortion types include GBLUR, FF, JP2K, JPEG, and WN. For each distortion type, a reference 3D image generates 3 symmetrically and 6 asymmetrically distorted images. The subjective evaluation values in the database are given in the form of DMOS, and the values of

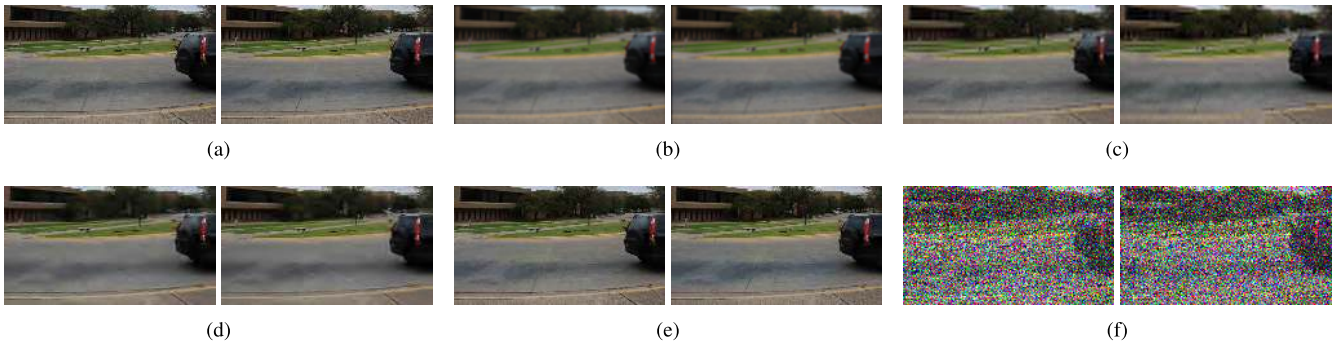

FIGURE 7. Sample images in the LIVE 3D IQA database Phase-I. (a) Reference. (b) GBLUR. (c) FF. (d) JP2K. (e) JEPG. (f) WN.

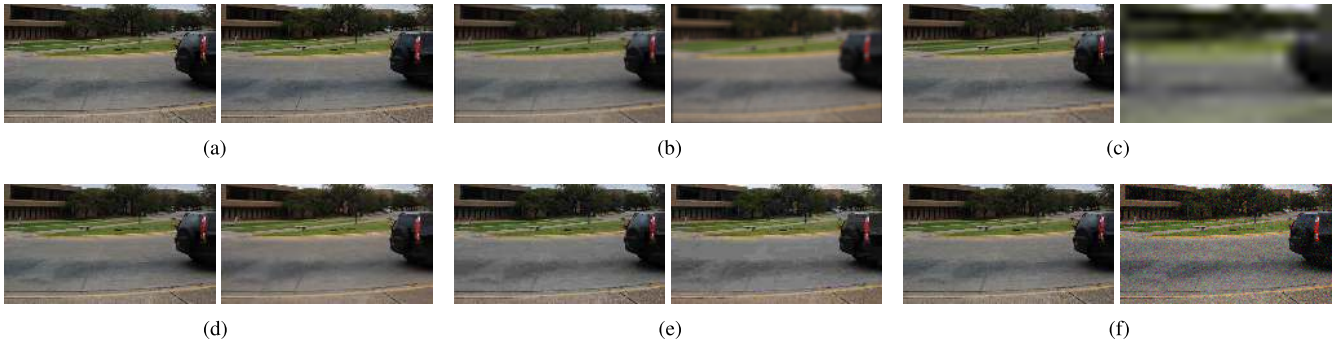

FIGURE 8. Sample images in the LIVE 3D IQA database Phase-II. (a) Reference. (b) Blur. (c) FF. (d) JP2K. (e) JEPG. (f) WN.

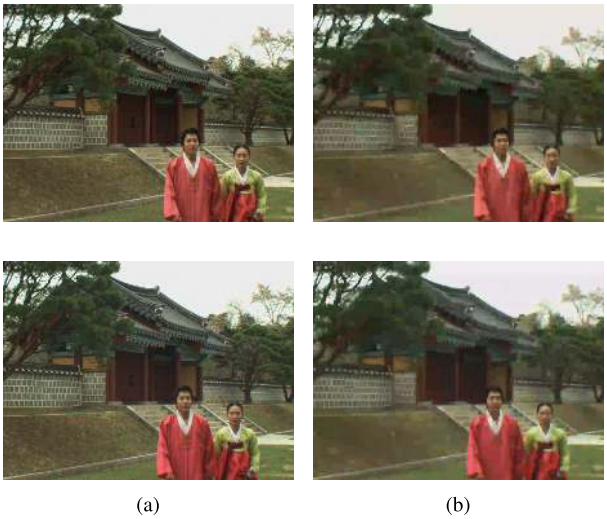

FIGURE 9. Sample images of the NBU 3D IQA database. (a) Reference. (b) H.264.

DMOS are distributed in the interval of [0, 100]. Fig. 8 shows a reference 3D image and its asymmetrically distorted images distorted by 5 distortion types.

3) NBU 3D IQA DATABASE [126]

It contains 12 reference 3D images and 312 symmetrically distorted images. The distortion types include JPEG, JP2K, GBULR, WN, and H.264. The subjective values for the distorted images are given by DMOS, and the values of DMOS are distributed in the interval of [0, 100]. Fig. 9 shows a reference 3D image and its an asymmetrically distorted image distorted by H.264.

In Table 4, we show the experimental results of 6 FR and 9 NR 3D image fidelity assessment metrics. Because the distortions introduced in the left and right views are the same for symmetrically distorted images and different for asymmetrically distorted images, the fidelity assessment for symmetrically distorted images is easier than that for asymmetrically distorted images. So the experimented metrics usually perform better on the LIVE 3D IQA database Phase-I and NBU 3D IQA database than on the LIVE 3D IQA database Phase-II. In recent years, machine learning and deep learning based metrics [71], [75], [76], [80], [81] have achieved significant performance improvements, especially for asymmetrically distorted images. The performance of state-or-the-art NR metrics [76], [80]–[82] is similar to that of state-or-the-art FR metrics [71], [72].

C. EXPERIMENTAL RESULTS ON IMAGE AESTHETICS ASSESSMENT

We first give the summarized information on commonly used image aesthetics assessment databases in Table 5. The detailed descriptions are given as follows.

1) AESTHETIC VISUAL ANALYSIS (AVA) DATABASE [127]

AVA database contains more than 250,000 images, each image contains a large number of aesthetic scores. In addition to the aesthetic evaluation, there are 66 semantic and 14 photographic style annotations. According to work [103], half of the images are labeled as high aesthetics, while the rest are

TABLE 4. Experimental results of existing 3D image fidelity assessment metrics on three databases. The top three performance values for FR and NR metrics are formatted in bold and highlighted in red, green, and blue colors, respectively. And the symbol "-" indicates that the value is not provided in the corresponding paper and we could not find the corresponding source code.

Type	Metric	LIVE 3D Phase-I			LIVE 3D Phase-II			NBU 3D		
		SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE
FR	Benoit [65]	0.889	0.903	7.061	0.744	0.762	7.490	0.881	0.876	8.286
	Chen [68]	0.916	**0.917**	**6.533**	**0.889**	**0.900**	**4.987**	**0.909**	**0.908**	**7.185**
	Niu [69]	**0.927**	0.855	8.389	0.745	0.610	8.929	-	-	-
	Bensalma [70]	0.875	0.887	7.559	0.751	0.770	7.204	**0.938**	**0.937**	**6.017**
	Shao [71]	**0.925**	**0.935**	**5.816**	**0.849**	**0.863**	**5.706**	**0.941**	**0.941**	**5.800**
	Wang [72]	**0.924**	**0.929**	**6.048**	**0.918**	**0.915**	**4.549**	-	-	-
NR	Ryu [73]	0.860	0.800	7.930	-	-	-	-	-	-
	Shao [74]	0.894	0.899	-	-	-	-	**0.853**	**0.862**	-
	Zhou [75]	**0.921**	**0.941**	**5.540**	**0.919**	**0.923**	**4.262**	-	-	-
	Shao [76]	**0.944**	**0.953**	-	**0.885**	0.903	-	**0.938**	**0.949**	-
	Shao [77]	0.867	0.885	-	0.872	0.909	-	-	-	-
	Chen [79]	0.891	0.895	7.247	0.880	0.895	5.102	-	-	-
	Jiang [80]	0.912	0.930	**6.031**	**0.915**	**0.922**	**4.362**	**0.931**	**0.936**	**6.019**
	Zhang [81]	**0.943**	**0.947**	**5.336**	-	-	-	-	-	-
	Shen [82]	-	-	-	**0.919**	**0.919**	**3.641**	-	-	-

FIGURE 10. Sample images in the AVA database. (a) Images in the green- and red-framed boxes are labeled with a mean score greater and smaller than 5, respectively. Images on the right in both green- and red-framed boxes are labeled with a mean score around five. (b) The number of images in the AVA database. (This figure comes from [130]).

TABLE 5. Summarized information of image aesthetics assessment databases.

Database	# of images	Year	Reference
AVA	250,000	2012	[127]
CUHK-PQ	17,690	2013	[93]
HB	15,000	2015	[128]
Kodak	1,500	2010	[129]
LIVE-IQ	779	2006	[124]

labeled as low aesthetics. Fig. 10 shows some sample images in the AVA database.

2) CUHK-PQ DATABASE [93]

It contains more than 17,690 images, half of which are labeled as high aesthetics and the rest are labeled as low aesthetics. To assign labels to the images, each image was viewed by 10 people, who labeled the image as high or low aesthetics. If at least 8 out of 10 people agreed with their assessment of the image, the image was kept and assigned a final label. Fig. 11 shows some sample images in the CUHK-PQ database.

3) HIDDEN BEAUTY OF FLICKR PICTURES (HB) DATABASE [128]

It contains more than 15,000 images which were chosen from a large YFCC100M database [131]. Each image falls into one of four categories: people, urban, nature, and animals. Each image was rated five times. The aesthetic score of an image is the average of all of its scores. Fig. 12 shows some sample images in the HB database.

4) KODAK AESTHETICS DATABASE (KODAK) [129]

It contains more than 1,500 images, and is an extended version of database described in Jiang *et al.* [129]. Each image was evaluated by four people on a scale of 1-100. The ground truth score of an image is mean of its four scores. Fig. 13 shows some sample images in the Kodak database.

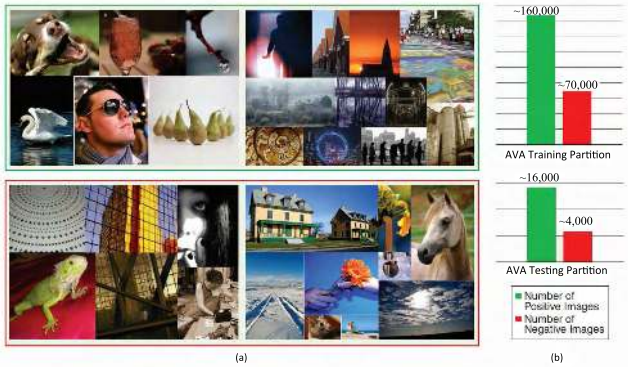

FIGURE 11. Some sample images in the CUHK-PQ database. (a) Images in the green- and red-framed boxes are labeled as high and low aesthetics, respectively. (b) The number of images in the CUHK-PQ database. (This figure comes from [130]).

FIGURE 12. Sample images in the HB database. (This figure comes from [128]).

FIGURE 13. Sample images in the Kodak database. (This figure comes from [129]).

5) LIVE IMAGE QUALITY (LIVE-IQ) DATABASE [124]

It contains 779 distorted images evaluated by nearly 25,000 individual human quality judgments. According to work [103], half of the images are labeled as high aesthetics, while the rest are labeled as low aesthetics. Fig. 14 shows some sample images in the LIVE-IQ database.

In Table 6, we show the experimental results of 8 image aesthetics assessment metrics: Nishiyama *et al.* [87], Marchesotti *et al.* [89], Karayev *et al.* [95], Lu *et al.* [96], Lu *et al.* [97], Mai *et al.* [98], Kong *et al.* [99], and Liu *et al.* [103], on CUHK-PQ, AVA, and LIVE-IQ databases. The performance indicator that is commonly adopted by image aesthetics assessment metrics on these three databases is accuracy. The experimental results for Nishiyama *et al.* [87], Marchesotti *et al.* [89], Lu *et al.* [96], Lu *et al.* [97], Mai *et al.* [98], and Liu *et al.* [103] are from Liu *et al.* [103]. As show in Table 6, among all experimented metrics, the metric presented by Liu *et al.* [103] performs best and achieves significant performance improvements on CUHK-PQ, AVA, and LIVE-IQ databases.

FIGURE 14. Sample images in the LIVE-IQ database. Images in the top and bottom rows are labeled as high and low aesthetics, respectively.

TABLE 6. Experimental results (accuracies) of existing image aesthetics assessment metrics on CUHK-PQ, AVA, and LIVE-IQ databases. The top three performance values for each databases are formatted in bold and highlighted in red, green, and blue colors, respectively. The symbol "-" indicates that the value is not provided in the corresponding paper and we could not find the corresponding source code.

Method	CUHK-PQ	AVA	LIVE-IQ
Nishiyama [88]	0.7745	0.7659	0.8657
Marchesotti [89]	0.8767	**0.7891**	0.8784
Karayev [95]	-	0.604	-
Lu [96]	0.9154	0.7446	0.8832
Lu [97]	**0.9237**	0.7446	**0.9023**
Mai [98]	**0.9276**	0.771	**0.8943**
Kong [99]	-	**0.7733**	-
Liu [103]	**0.9341**	**0.8309**	**0.9465**

TABLE 7. Experimental results (correlation coefficients) of existing image aesthetics assessment metrics on HB and Kodak databases. The top three performance values for each databases are formatted in bold and highlighted in red, green, and blue colors, respectively.

Method	HB	Kodak
Datta [83]	**0.413**	**0.571**
Ke [84]	**0.431**	**0.547**
Luo [85]	0.296	0.293
Marchesotti [89]	0.258	**0.310**
Tang [93]	**0.458**	0.297

TABLE 8. Summarized information of three 3D image depth VCA databases.

Database	# of img.	Resolution	Year	Ref.
NBU 3D-VCA	200	1920x1080	2015	[113]
IVY Lab 3D	120	1920x1080	2012	[132]
IEEE SA	800	1920x1080	2012	[117]

In Table 7, we show the experimental results of 5 image aesthetics assessment metrics: Datta *et al.* [83], Ke *et al.* [84], Luo and Tang [85], Marchesotti *et al.* [89], and Luo *et al.* [93], on HB and Kodak databases. The performance indicator that is commonly adopted by image aesthetics assessment metrics on HB and Kodak databases is the correlation coefficient [102]. As show in Table 7, among all experimented metrics, the metric presented by Luo *et al.* [93] performs best on HB database, and the metric presented by Datta *et al.* [83] performs best on Kodak Database.

FIGURE 15. Sample images in the NBU 3D-VCA Image Database.

FIGURE 16. Sample images in the IVY Lab Stereoscopic 3D Image Database.

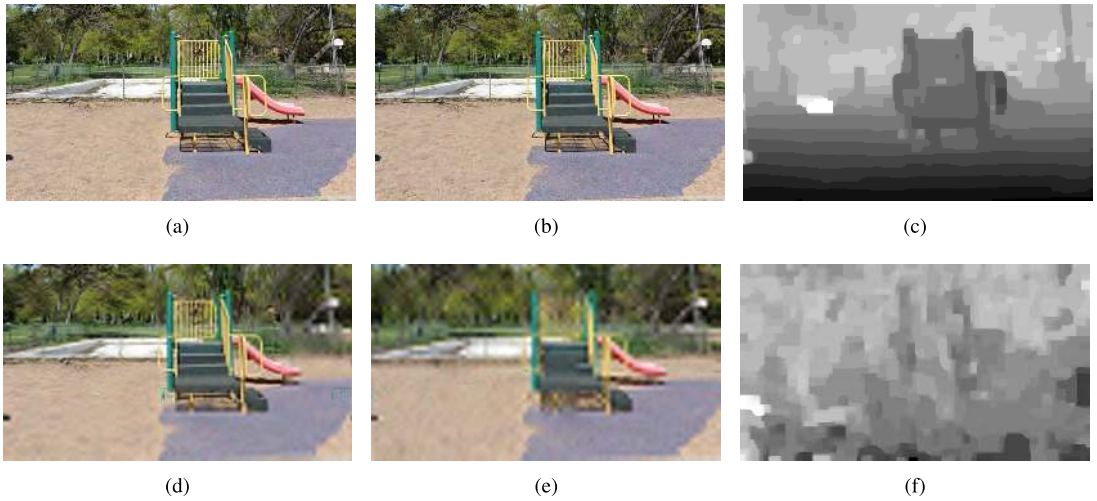

FIGURE 17. Example of a reference 3D image (a and b), its symmetrically distorted 3D image (c and d), and the corresponding disparity maps (c and f). (a) Left reference view. (b) Right reference view. (c) Disparity map of (a) and (b). (d) Left distored view. (e) Right distored view. (f) Disparity map of (d) and (e).

D. EXPERIMENTAL RESULTS ON 3D IMAGE VISUAL COMFORT ASSESSMENT

There are three commonly used 3D image depth VCA databases: NBU 3D-VCA Image Database, IVY Lab Stereoscopic 3-D Image Database, and IEEE Standards Association Stereoscopic Database (IEEE SA). Because calculating the values of MOS is expensive, Jiang *et al.* [113] proposed using the preference label to replace the MOS, and then created the NBU 3D-VCA database for further 3D image VCA researches. Table 8 shows the summarized information of these three 3D image depth VCA databases.

1) NBU 3D-VCA IMAGE DATABASE [113]

This database contains a total of 200 3D images with the resolution of 1920×1080 pixels. The images in this database are all captured by a Sony HDRTD30E dual-lens 3D camera. These images include 82 indoor and 118 outdoor scenes with a variety of textures, colors, and depth ranges. To reflect a large variety of visual comfort, the maximum ranges of negative disparity in these images range from 0.02 to 4.79 deg. Some sample images in the database are shown in Fig. 15.

2) IVY LAB STEREOSCOPIC 3D IMAGE DATABASE [132]

This database contains 120 3D images consisting of 62 indoor and 58 outdoor scenes (see Fig. 16). These 3D images include diverse scenes, such as humans, trees, structures, and man-made objects. The images were captured using a hand-held 3D digital camera with a resolution of 1920×1080 pixels. The maximum disparities of the 3D images are distributed in the range from 0 to 5 deg.

3) IEEE SA DATABASE [117]

It contains 800 3D images with a resolution of 1920×1080 pixels. For each scene, there are multiple 3D images corresponding to multiple evenly separated convergence points. There are 160 such convergence-sampled sets in the IEEE SA database.

In Table 9, we show the experimental results of 10 3D image depth VCA metrics on these databases. The performance indicators that are commonly adopted by image depth VCA metrics are PLCC, SRCC, and RMSE.

As shown in Table 9, most depth VCA metrics validate their performance on a single database. For the NBU

TABLE 9. Experimental results of different depth VCA metrics on three databases. The best three performance values on each database are formatted in bold and highlighted in red, green, blue colors, respectively. And the symbol "-" indicates that the value is not provided in the corresponding paper and we could not find the corresponding source code.

Metric	NBU 3D-VCA			IVY Lab			IEEE SA		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
Zhou [115]	-	-	-	**0.865**	-	**0.416**	-	-	-
Jung [117]				**0.849**	**0.811**	0.440			
Jiang [113]	0.8078	0.7682	0.4623	**0.8369**	**0.8210**	**0.4377**	-	-	-
Jiang [114]	**0.8351**	**0.7745**	**0.4409**						
Jiang [119]	**0.8460**	**0.7889**	**0.4318**	-	-	-	-	-	-
Sohn [132]	-	-	-	-	-	**0.422**	-	-	-
Park [111]	-	-	-	-	-	-	0.8310	0.7534	**0.4489**
Jeong [112]	-	-	-	-	-	-	**0.900**	**0.825**	**0.351**
Ahn [116]	-	-	-	-	-	-	**0.8512**	**0.8647**	-
Kim [120]	**0.8128**	**0.7684**	**0.4015**	-	-	-	**0.9402**	**0.8432**	**0.3491**

FIGURE 18. Sample images in the NVIDIA 3D VISION LIVE Highest Rated database.

3D-VCA database, the best performance is achieved by the metric presented by Jiang *et al.* [119]. For the IVY Lab and IEEE SA databases, the best performance is achieved by the metrics presented by Zhou *et al.* [115] and Kim *et al.* [120], respectively.

IV. CHALLENGES

In this section, we present three challenges for the state-of-the-art IQA metrics.

A. INFLUENCE OF DIFFERENT FACTORS ON EACH OTHER

Existing research on image quality assessment evaluates image quality in isolation, such as from a single perspective of distortion, aesthetics, or visual comfort. However, distortion, aesthetics, and visual comfort influence each other in IQA. As shown in Fig. 17, the distortions in the left and right views of a 3D image also changes the disparity map (Fig. 17 (f)) which determines the visual comfort of the 3D image.

Distortion also affects aesthetics assessment. Some photographic rules detection algorithms have been presented to identify whether or not a photo respects some photographic rules, such as the rule-of-thirds [133] and the rule of simplicity [134]. Because these algorithms extract visual features from saliency maps which are usually affected by distortions in images [2], [3], the distortion usually reduces the performance of these aesthetics assessment metrics.

To summarize, the influence of distortion, aesthetics, and visual comfort on each other are not well addressed in existing IQA metrics.

B. PERFORMANCE IN REAL APPLICATIONS

Existing works assess the quality of an image just from one or two perspectives and use databases with subjective opinion scores obtained in lab experiments. However, when people evaluate the image quality, they integrate various influencing factors and use them as cues to comprehensively evaluate the image quality. Therefore, the consistency between existing image quality assessment metrics and the subjective quality assessment scores in real applications of quality assessment is questionable.

We experimented with a variety of quality assessment metrics on a database from a real application called NVIDIA 3D VISION LIVE Highest Rated[1] images. Unlike the databases described in Section 3, this database was collected from a website and is a database from a real application. The visitors of the website who voted were not constrained to any perspectives, and hence the votes reflect their comprehensive subjective opinions. In the database, the 3D images cover a variety of classes, including people, animals, plants, landscapes, and cityscapes. The average subjective scores range from 1 to 5. A large score indicates a good quality image. Because the numbers of votes for different images were different, we only selected the images that had at least three votes, which were 730 3D images. Fig. 18 shows some images from this database.

We experimented with 8 state-of-the-art metrics, including Jiang *et al.* [113], Kong *et al.* [99], Mai *et al.* [98], One-column CNN [81], NSS [51], GWH-GLBP [135], dipIQ [136], and DIQaM-NR [40], on this database. Among all

[1] http://photos.3dvisionlive.com/dn/highest-rated/

TABLE 10. Experimental results of different IQA metrics and their combinations on the NVIDIA database. The best three performance values are formatted in bold and highlighted in red, green and blue colors, respectively.

	PLCC	SRCC	RMSE
Jiang [113]	0.7910	0.7684	0.5709
one-column CNN [81]	0.5981	0.6021	0.6940
NSS [51]	0.4452	0.4231	0.7114
GWH-GLBP [135]	0.4800	0.4590	0.5568
dipIQ [136]	0.5007	0.4264	0.6067
DIQaM-NR [40]	0.5094	0.4805	0.6289
Kong [99]	0.1113	0.0430	0.5841
Mai [98]	0.1778	0.1239	0.5771
Jiang+ one-column CNN	0.8040	0.7801	0.5665
Jiang+NSS	0.8271	0.7996	0.5609
Jiang+GWH-GLBP	0.8256	**0.8300**	0.5428
Jiang+dipIQ	0.8301	0.7901	**0.5291**
Jiang+DIQaM-NR	**0.8357**	0.8150	0.5403
Jiang+Kong	0.7986	0.7697	0.5686
Jiang+Mai	0.7924	0.7586	0.5679
ALL except [98] and [99]	**0.8434**	**0.8311**	**0.5402**
ALL	**0.8435**	**0.8342**	**0.5147**

these metrics, Jiang *et al.* [113] focuses on visual comfort, Kong *et al.* [99] and Mai *et al.* [98] focus on aesthetics, and the other metrics focus on distortion. Because the authors' implementation for Jiang *et al.* [113] is not available, we implemented and used the disparity features that the authors extracted in the comparison experiments. To adapt NSS [51], GWH-GLBP [135], dipIQ [136], and DIQaM-NR [40] to 3D images, we used their methods to extract the features from both the left and right views, respectively. We used the pre-trained models for dipIQ [136] and DIQaM-NR [40] without fine-tuning. Because the pre-trained model for one-column CNN [81] was not available, we implemented the CNN and trained a model using the NVIDIA 3D VISION LIVE Highest Rated database. We randomly chosen 70% images in the database as the training images and the left 30% images were used as the testing images. The random forest regression method was used as the machine learning method. The experiments were repeated 100 times, and we report the average values of PLCC, SRCC, and RMSE in the paper. The experimental results of these metrics are shown in Table 10.

As shown in Table 10, the Jiang *et al.* [113] index achieved the best performance among the experimented 8 metrics. Kong *et al.* [99] and Mai *et al.* [98] achieved lower performance than other metrics. We also combined Jiang *et al.* [113] with each of the distortion and aesthetics related quality assessment metrics. Besides, we combined Jiang *et al.* [113] with all distortion related quality assessment metrics and showed the experimental results in the second row from the bottom. Finally, we combined Jiang *et al.* [113] with all distortion and aesthetics related quality assessment metrics, and the experimental results were shown in the last row. The superior performance of the combinations indicates that all the different factors matter for quality assessment, and that these factors should be considered together to improve the performance of quality assessment in real applications.

C. COMBINATION OF QUALITY ASSESSMENT, RESTORATION, AND ENHANCEMENT

Some quality assessment metrics [38], [51], [137]–[139] also predict the distortion types of the distorted images. This information can be used to guide quality restoration and enhancement. For example, images with blur, noise, and compression can use the corresponding deblurring [7], denosing [8], [140], and deblocking [9] algorithms. However these quality enhancement algorithms are presented for specific distortions and need the guidance from quality assessment metrics.

Yu *et al.* [141] presented a toolchain for image restoration by deep reinforcement learning (RL-Restore). In comparison to the above mentioned image quality enhancement algorithms, RL-Restore learns to select an appropriate tool from the toolbox to iteratively enhance the quality of a distorted image. RL-Restore can enhance the quality of an image distorted with multiple and unknown distortions. However RL-Restore can only handle limited types of distortions, including blur, noise, and JPEG compression.

Enhancing the quality of an image distorted by other types of distortions, poor aesthetics, and low visual comfort is still challenging for the combination of image quality assessment, restoration, and enhancement.

V. CONCLUSIONS

In this paper, we first gave an up-to-date overview on IQA from the perspectives of fidelity, aesthetics, and visual comfort. We then described the IQA databases and the experimental results of representative IQA metrics. Image fidelity assessment has received more attention from the researchers as compared to image aesthetics assessment and visual comfort assessment.

We also presented three challenges for IQA. First, the influence of distortion, aesthetics, and visual comfort on image quality are not well addressed by existing IQA metrics. Second, we showed that the performance of existing IQA metrics in real applications needs to be improved. Finally, we showed that the combination of quality assessment, restoration, and enhancement needs to handle more quality problems, including all types of distortions, poor aesthetics, and low visual comfort.

REFERENCES

[1] P. Korshunov and W. T. Ooi, "Video quality for face detection, recognition, and tracking," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 7, no. 3, 2011, Art. no. 14.

[2] Y. Niu, L. Ke, and W. Guo, "Evaluation of visual saliency analysis algorithms in noisy images," *Mach. Vis. Appl.*, vol. 27, no. 6, pp. 915–927, 2016.

[3] Y. Niu, L. Lin, Y. Chen, and L. Ke, "Machine learning-based framework for saliency detection in distorted images," *Multimedia Tools Appl.*, vol. 76, no. 24, pp. 26329–26353, 2017.

[4] E. Kafetzakis, C. Xilouris, M. A. Kourtis, M. Nieto, I. Jargalsaikhan, and S. Little, "The impact of video transcoding parameters on event detection for surveillance systems," in *Proc. IEEE Int. Symp. Multimedia*, Dec. 2013, pp. 333–338.
[5] Y. Fang, Y. Yuan, L. Li, J. Wu, W. Lin, and Z. Li, "Performance evaluation of visual tracking algorithms on video sequences with quality degradation," *IEEE Access*, vol. 5, pp. 2430–2441, 2017.
[6] K.-Y. Lin and G. Wang, "Hallucinated-IQA: No-reference image quality assessment via adversarial learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2018, pp. 723–731.
[7] L. Mai and F. Liu, "Kernel fusion for better image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 371–380.
[8] P. Jain and V. Tyagi, "LAPB: Locally adaptive patch-based wavelet domain edge-preserving image denoising," *Inf. Sci.*, vol. 294, no. 10, pp. 164–181, 2015.
[9] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu, "Efficient deblocking with coefficient regularization, shape-adaptive filtering, and quantization constraint," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 735–745, Aug. 2008.
[10] M. Beech, *Digital 3D Stereo Guide*. New York, NY, USA: Self Publishing, 2009.
[11] B. Mendiburu, *3D Movie Making: Stereoscopic Digital Cinema From Script to Screen*. Waltham, MA, USA: Focal Press, 2009.
[12] Y. Niu, H. Zhang, W. Guo, and R. Ji, "Image quality assessment for color correction based on color contrast similarity and color value difference," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 4, pp. 849–862, Apr. 2018.
[13] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
[14] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS," in *Proc. Int. Workshop Video Process. Qual. Metrics*, vol. 4, 2006, pp. 1–4.
[15] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, and M. Carli, "Modified image visual quality metrics for contrast change and mean shift accounting," in *Proc. Int. Conf. Exper. Designing Appl. CAD Syst. Microelectron. (CADSM)*, 2011, pp. 305–311.
[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[17] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[18] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2004, pp. 1398–1402.
[19] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2011, pp. IV-3313–IV-3316.
[20] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2005, pp. 573–576.
[21] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
[22] X. Zhang, X. Feng, W. Wang, and W. Xue, "Edge strength similarity for image quality assessment," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 319–322, Apr. 2013.
[23] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 520–531, Sep. 2015.
[24] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011006-1–011006-21, 2010.
[25] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[26] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
[27] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Aug. 2014.
[28] Y. Niu, J. Chen, and W. Guo, "Meta-metric for saliency detection evaluation metrics based on application preference," *Multimedia Tools Appl.*, vol. 77, no. 20, pp. 26351–26369, 2018.
[29] Y. Niu, W. Lin, X. Ke, and L. Ke, "Fitting-based optimisation for image visual salient object detection," *IET Comput. Vis.*, vol. 11, no. 2, pp. 161–172, 2017.
[30] Y. Niu, W. Lin, and X. Ke, "CF-based optimisation for saliency detection," *IET Comput. Vis.*, vol. 12, no. 4, pp. 365–376, Jun. 2018.
[31] W. Guo, X. Sun, and Y. Niu, "Multi-scale saliency detection via inter-regional shortest colour path," *IET Comput. Vis.*, vol. 9, no. 2, pp. 290–299, 2015.
[32] A. Balanov, A. Schwartz, Y. Moshe, and N. Peleg, "Image quality assessment based on DCT subband similarity," in *Proc. IEEE Int. Conf. Image Process.*, vol. 27, no. 5, Sep. 2015, pp. 311–325.
[33] H. Z. Nafchi, A. Shahkolaei, R. Hedjam, and M. Cheriet, "Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator," *IEEE Access*, vol. 4, pp. 5579–5590, 2017.
[34] W. Sun, Q. Liao, J.-H. Xue, and F. Zhou, "SPSIM: A superpixel-based similarity index for full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4232–4244, Sep. 2018.
[35] M. Narwaria and W. Lin, "SVD-based quality metric for image and video using machine learning," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 42, no. 2, pp. 347–364, Apr. 2012.
[36] C. Charrier, O. Lézoray, and G. Lebrun, "Machine learning to design full-reference image quality assessment algorithm," *Signal Process., Image Commun.*, vol. 27, no. 3, pp. 209–219, 2012.
[37] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT 500–11, 2002.
[38] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1793–1807, May 2013.
[39] S. Wang, C. Deng, W. Lin, G. Huang, and B. Zhao, "NMF-based image quality assessment using extreme learning machine," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 232–243, Jan. 2017.
[40] S. Bosse, D. Maniry, K. R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
[41] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1969–1977.
[42] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824–829, Aug. 2011.
[43] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.
[44] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Inf. Sci.*, vol. 351, pp. 18–29, Jul. 2016.
[45] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2013, pp. 1095–1098.
[46] K. Gu, G. T. Zhai, and M. Lin, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2015.
[47] S. Wang, C. Deng, B. Zhao, G.-B. Huang, and B. Wang, "Gradient-based no-reference image blur assessment using extreme learning machine," *Neurocomputing*, vol. 174, pp. 310–321, Jan. 2015.
[48] Y. Zhan and R. Zhang, "No-reference JPEG image quality assessment based on blockiness and luminance change," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 760–764, Jun. 2017.
[49] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, and B. Chen, "No-reference quality assessment of deblocked images," *Neurocomputing*, vol. 177, pp. 572–584, Feb. 2016.
[50] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, Jul. 2015.
[51] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[52] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[53] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[54] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 2013–2026, Dec. 2013.

[55] Q. Wang, J. Chu, L. Xu, and Q. Chen, "A new blind image quality framework based on natural color statistic," *Neurocomputing*, vol. 173, pp. 1798–1810, Jan. 2016.

[56] Y. Zhang, J. Wu, X. Xie, L. Li, and G. Shi, "Blind image quality assessment with improved natural scene statistics model," *Digit. Signal Process.*, vol. 57, pp. 56–65, Oct. 2016.

[57] Y. Li *et al.*, "No-reference image quality assessment with shearlet transform and deep neural networks," *Neurocomputing*, vol. 154, pp. 94–109, Apr. 2015.

[58] W. Zhou, L. Yu, W. Qiu, Y. Zhou, and M.-W. Wu, "Local gradient patterns (LGP): An effective local-statistical-feature extraction scheme for no-reference image quality assessment," *Inf. Sci.*, vols. 397–398, pp. 1–14, Aug. 2017.

[59] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1733–1740.

[60] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.

[61] D. Pan, P. Shi, M. Hou, Z. Ying, S. Fu, and Y. Zhang, "Blind predicting similar quality map for image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, May 2018, pp. 6373–6382.

[62] K.-Y. Lin and G. Wang, "Hallucinated-IQA: No-reference image quality assessment via adversarial learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2018, pp. 732–741.

[63] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, "Quality analysis for 3D video using 2D video quality models," *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1969–1976, Nov. 2008.

[64] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE*, vol. 6803, pp. 680305-1–680305-12, Feb. 2008.

[65] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, no. 1, pp. 659024-1–659024-13, 2009.

[66] M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *Proc. Int. Conf. Image Process.*, Sep. 2003, pp. 1–4.

[67] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. Int. Workshop Video Process. Qual. Metrics Consum. Electron.*, 2010, pp. 1–6.

[68] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Process. Image Commun.*, vol. 28, no. 9, pp. 1143–1155, 2013.

[69] J. Zhan, Y. Niu, and Y. Huang, "Learning from multi metrics for stereoscopic 3D image quality assessment," in *Proc. Int. Conf. 3D Imag.*, Dec. 2017, pp. 1–8.

[70] R. Bensalma and M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Syst. Signal Process.*, vol. 24, no. 2, pp. 281–316, 2013.

[71] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, and Q. Dai, "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 2971–2983, Oct. 2015.

[72] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.

[73] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 591–602, Apr. 2014.

[74] F. Shao, W. Lin, S. Wang, G. Jiang, and M. Yu, "Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 154–165, Jun. 2015.

[75] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.-W. Wu, and T. Luo, "Blind quality estimator for 3D images based on binocular combination and extreme learning machine," *Pattern Recognit.*, vol. 71, pp. 207–217, Nov. 2017.

[76] F. Shao, K. Li, W. Lin, G. Jiang, and Q. Dai, "Learning blind quality evaluator for stereoscopic images using joint sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2104–2114, Oct. 2016.

[77] F. Shao, Z. Zhang, Q. Jiang, W. Lin, and G. Jiang, "Toward domain transfer for no-reference quality prediction of asymmetrically distorted stereoscopic images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 573–585, Mar. 2018.

[78] R. Akhter, Z. M. P. Sazzad, Y. Horita, and J. Baltes, "No-reference stereoscopic image quality assessment," *Proc. SPIE*, vol. 7524, pp. 75240T-1–75240T-12, Feb. 2010.

[79] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3379–3391, Sep. 2013.

[80] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "Learning a referenceless stereopair quality engine with deep nonnegativity constrained sparse autoencoder," *Pattern Recognit.*, vol. 76, pp. 242–255, Apr. 2018.

[81] W. Zhang, C. Qu, L. Ma, J. Guan, and R. Huang, "Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network," *Pattern Recognit.*, vol. 59, pp. 176–187, Nov. 2016.

[82] L. Shen, J. Lei, and C. Hou, "No-reference stereoscopic 3D image quality assessment via combined model," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8195–8212, 2018.

[83] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 288–301.

[84] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 419–426.

[85] Y. Luo and X. Tang, "Photo and video quality evaluation: Focusing on the subject," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 386–399.

[86] S. Bhattacharya, R. Sukthankar, and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in *Proc. ACM Int. Conf. Multimedia*, 2010, pp. 271–280.

[87] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato, "Sensation-based photo cropping," in *Proc. ACM Int. Conf. Multimedia*, 2009, pp. 669–672.

[88] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato, "Aesthetic quality classification of photographs based on color harmony," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 33–40.

[89] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1784–1791.

[90] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[91] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the Fisher kernel for large-scale image classification," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 143–156.

[92] H.-H. Su, T.-W. Chen, C.-C. Kao, W. H. Hsu, and S.-Y. Chien, "Scenic photo quality assessment with bag of aesthetics-preserving features," in *Proc. ACM Int. Conf. Multimedia*, 2011, pp. 1213–1216.

[93] X. Tang, W. Luo, and X. Wang, "Content-based photo quality assessment," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1930–1943, Dec. 2013.

[94] L. Marchesotti, N. Murray, and F. Perronnin, "Discovering beautiful attributes for aesthetic image analysis," *Int. J. Comput. Vis.*, vol. 113, no. 3, pp. 246–266, Jul. 2014.

[95] S. Karayev *et al.*, "Recognizing image style," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 1–20.

[96] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "RAPID: Rating pictorial aesthetics using deep learning," in *Proc. ACM Int. Conf. Multimedia*, 2014, pp. 457–466.

[97] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 990–998.

[98] L. Mai, H. Jin, and F. Liu, "Composition-preserving deep photo aesthetics assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 497–506.

[99] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes, "Photo aesthetics ranking network with attributes and content adaptation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 662–679.

[100] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 1097–1105, 2012.
[101] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a 'Siamese' time delay neural network," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 7, no. 4, pp. 669–688, 1993.
[102] M. Kucer, A. C. Loui, and D. W. Messinger, "Leveraging expert feature knowledge for predicting image aesthetics," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5100–5112, Oct. 2018.
[103] Z. Liu, Z. Wang, Y. Yao, L. Zhang, and L. Shao, "Deep active learning with contaminated tags for image aesthetics assessment," *IEEE Trans. Image Process.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8340874
[104] Y. Niu, F. Liu, W.-C. Feng, and H. Jin, "Aesthetics-based stereoscopic photo cropping for heterogeneous displays," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 783–796, Jun. 2012.
[105] M. B. Islam, W. Lai-Kuan, W. Chee-Onn, and K.-L. Low, "Stereoscopic image warping for enhancing composition aesthetics," in *Proc. Asian Conf. Pattern Recognit.*, Nov. 2016, pp. 645–649.
[106] M. B. Islam, L.-K. Wong, K.-L. Low, and C.-O. Wong, "Aesthetics-driven stereoscopic 3-D image recomposition with depth adaptation," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 2964–2979, Nov. 2018.
[107] F. Shao, W. Lin, Z. Li, G. Jiang, and Q. Dai, "Toward simultaneous visual comfort and depth sensation optimization for stereoscopic 3-D experience," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4521–4533, Dec. 2017.
[108] W. A. IJsselsteijn, H. D. Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 225–233, Mar. 2000.
[109] P. J. H. Seuntiëns, "Visual experience of 3D TV," M.S. thesis, Eindhoven Univ. Technol., Eindhoven 2006. [Online]. Available: https://pure.tue.nl/ws/files/1687707/200610884.pdf, doi: 10.6100/IR609714.
[110] M. Bernhard, C. Dell'mour, M. Hecher, E. Stavrakis, and M. Wimmer, "The effects of fast disparity adjustment in gaze-controlled stereoscopic applications," in *Proc. Symp. Eye Tracking Res. Appl.*, 2014, pp. 111–118.
[111] J. Park, H. Oh, S. Lee, and A. C. Bovik, "3D visual discomfort predictor: Analysis of disparity and neural activity statistics," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1101–1114, Mar. 2015.
[112] H. Jeong, H. G. Kim, and Y. M. Ro, "Visual comfort assessment of stereoscopic images using deep visual and disparity features based on human attention," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 715–719.
[113] Q. Jiang, F. Shao, G. Jiang, M. Yu, and Z. Peng, "Three-dimensional visual comfort assessment via preference learning," *J. Electron. Imag.*, vol. 24, no. 4, pp. 043002-1–043002-12, 2015.
[114] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "On predicting visual comfort of stereoscopic images: A learning to rank based approach," *IEEE Signal Process. Lett.*, vol. 23, no. 2, pp. 302–306, Feb. 2016.
[115] Y. Zhou, Y. He, S. Zhang, and Y. Zhang, "Visual comfort prediction for stereoscopic image using stereoscopic visual saliency," *Multimedia Tools Appl.*, vol. 76, no. 22, pp. 23499–23516, 2017.
[116] S. Ahn, J. Kim, H. Kim, and S. Lee, "Visual attention analysis on stereoscopic images for subjective discomfort evaluation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2016, pp. 1–6.
[117] Y. J. Jung, H. Sohn, S.-I. Lee, H. W. Park, and Y. M. Ro, "Predicting visual discomfort of stereoscopic images using human attention model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 12, pp. 2077–2082, Dec. 2013.
[118] X. Zhang, J. Zhou, J. Chen, X. Guo, Y. Zhang, and X. Gu, "Visual comfort assessment of stereoscopic images with multiple salient objects," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, Jun. 2015, pp. 1–6.
[119] Q. Jiang, F. Shao, G. Jiang, M. Yu, and Z. Peng, "Leveraging visual attention and neural activity for stereoscopic 3D visual comfort assessment," *Multimedia Tools Appl.*, vol. 76, no. 7, pp. 9405–9425, 2016.
[120] H. G. Kim, H. Jeong, H.-T. Lim, and Y. M. Ro, "Binocular fusion net: Deep learning visual comfort assessment for stereoscopic 3D," *IEEE Trans. Circuits Syst. Video Technol.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8319944
[121] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2 (2005)," Lab. Image Video Eng., Univ. Texas, Austin, TX, USA, 2005. [Online]. Available: http://live.ece.utexas.edu/research/Quality/subjective.htm
[122] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Mod. Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.
[123] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. Eur. Workshop Vis. Inf. Process.*, 2013, pp. 106–111.
[124] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[125] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, vol. 28, no. 8, pp. 870–883, Dec. 2013.
[126] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.
[127] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 157, no. 10, Jun. 2012, pp. 2408–2415.
[128] R. Schifanella, M. Redi, and L. Aiello, "An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures," in *Proc. Int. Conf. Web Social Media*, 2015, pp. 296–317.
[129] W. Jiang, A. C. Loui, and C. D. Cerosaletti, "Automatic aesthetic value assessment in photographic images," in *Proc. IEEE Int. Conf. Multimedia Expo*, vol. 41, no. 3, Jul. 2010, pp. 920–925.
[130] Y. Deng, C. C. Loy, and X. Tang, "Image aesthetic assessment: An experimental survey," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 80–106, Jul. 2017.
[131] B. Thomee *et al.*, "YFCC100M: The new data in multimedia research," *Commun. ACM*, vol. 59, no. 2, pp. 4–73, Feb. 2016.
[132] H. Sohn, Y. J. Jung, S.-I. Lee, and Y. M. Ro, "Predicting visual discomfort using object size and disparity information in stereoscopic images," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 28–37, Mar. 2013.
[133] L. Mai, H. Le, Y. Niu, and F. Liu, "Rule of thirds detection from photograph," in *Proc. IEEE Int. Symp. Multimedia*, Dec. 2011, pp. 91–96.
[134] L. Mai, H. Le, Y. Niu, Y.-C. Lai, and F. Liu, "Detecting rule of simplicity from photos," in *Proc. ACM Int. Conf. Multimedia*, 2012, pp. 1149–1152.
[135] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.
[136] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, Aug. 2017.
[137] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
[138] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.
[139] L. Kang, P. Ye, Y. Li, and D. Doermann, "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 2791–2795.
[140] Y. Niu, Y. Yang, W. Guo, and L. Lin, "Region-aware image denoising by exploring parameter preference," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2433–2438, Sep. 2018.
[141] K. Yu, C. Dong, L. Lin, and C. C. Loy, "Crafting a toolchain for image restoration by deep reinforcement learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Nov. 2018, pp. 2443–2452.

YUZHEN NIU received the Ph.D. degree in computer science from Shandong University, China, in 2010. She was a Postdoctoral Researcher with the Department of Computer Science, Portland State University, Portland, OR, USA. She is currently a Professor with the College of Mathematics and Computer Science, Fuzhou University, China. Her current research interests include image processing, computer vision, artificial intelligence, and multimedia.

YINI ZHONG is currently pursuing the M.S. degree with the College of Mathematics and Computer Science, Fuzhou University, China. Her current research interests include computer vision and image processing.

WENZHONG GUO received the B.S. and M.S. degrees in computer science and the Ph.D. degree in communication and information systems from Fuzhou University, Fuzhou, China, in 2000, 2003, and 2010, respectively. He is currently a Full Professor with the College of Mathematics and Computer Science, Fuzhou University. He leads the Network Computing and Intelligent Information Processing Laboratory, which is a key Lab of Fujian Province, China. His research interests include mobile computing and evolutionary computation.

YIQING SHI received the B.S. degree in electronic information engineering from Fuzhou University, Fuzhou, China, in 2016, where she is currently pursuing the Ph.D. degree in communication and information system. Her current research interests include pattern recognition, image processing, and computer vision.

PEIKUN CHEN is currently pursuing the M.S. degree with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou, China. His current research interests include image processing, computer vision, and artificial intelligence.