SCIENCE CHINA
Information Sciences

• REVIEW •

November 2020, Vol. 63 211301:1–211301:52
https://doi.org/10.1007/s11432-019-2757-1

Perceptual image quality assessment: a survey

Guangtao ZHAI* & Xiongkuo MIN

Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, *China*

Received 3 August 2019/Revised 31 October 2019/Accepted 13 January 2020/Published online 26 April 2020

Abstract Perceptual quality assessment plays a vital role in the visual communication systems owing to the existence of quality degradations introduced in various stages of visual signal acquisition, compression, transmission and display. Quality assessment for visual signals can be performed subjectively and objectively, and objective quality assessment is usually preferred owing to its high efficiency and easy deployment. A large number of subjective and objective visual quality assessment studies have been conducted during recent years. In this survey, we give an up-to-date and comprehensive review of these studies. Specifically, the frequently used subjective image quality assessment databases are first reviewed, as they serve as the validation set for the objective measures. Second, the objective image quality assessment measures are classified and reviewed according to the applications and the methodologies utilized in the quality measures. Third, the performances of the state-of-the-art quality measures for visual signals are compared with an introduction of the evaluation protocols. This survey provides a general overview of classical algorithms and recent progresses in the field of perceptual image quality assessment.

Keywords visual quality assessment, human visual system, subjective quality assessment, objective quality assessment

Citation Zhai G T, Min X K. Perceptual image quality assessment: a survey. Sci China Inf Sci, 2020, 63(11): 211301, https://doi.org/10.1007/s11432-019-2757-1

1 Introduction

It is estimated that 1.2 trillion digital images were taken in 2017, thanks to the popularization of smartphones. In almost every stage of the visual communication systems, e.g., acquisition, compression, transmission, and display, various types of distortions are introduced. Quality assessment is needed to ensure and improve the quality of visual contents delivered to the end-users. Quality assessment metrics can be used as testing criteria or optimization goals for the visual communication systems. Quality assessment methods can be categorized into subjective and objective ones. Subjective assessment is usually considered as the most reliable and accurate because the human visual system (HVS) is the ultimate receiver of visual signals in most visual communication systems. However, subjective test is time-consuming and expensive, and cannot be directly embedded into a practical system as the optimization metric. Objective quality assessment methods, usually designed and/or trained using subjective assessment data, can predict the visual quality automatically, and are ideal for timely system performance evaluation and optimization.

Image and video are two types of visual contents to be transmitted by visual communication systems. In this survey, we focus on image quality assessment (IQA) owing to the following two reasons: first, a majority of research work is about images; second, image quality metrics usually serve as bases for video quality assessment (VQA) considering that videos are sequences of images.

* Corresponding author (email: zhaiguangtao@sjtu.edu.cn)

1.1 Related surveys

Over the past two decades, with the repaid flourishing of visual media, a large number of IQA metrics have been proposed, which call for a survey of the field. In 2009, Wang and Bovik [1] gave an analysis of full-reference (FR) image fidelity measure, with a pivot on mean square error (MSE). Wang and Bovik [2] later gave a more general introduction to reduced-reference (RR) and no-reference (NR) IQA, but only a limited number of metrics were reviewed. Lin and Kuo [3] gave a survey on perceptual visual quality metrics in 2011. They discussed several key problems related to IQA, including signal decomposition, just-noticeable distortion, visual attention, feature and artifact detection, feature pooling, viewing condition, computer-generated signal, and visual attention.

In 2011, Moorthy and Bovik [4] shared their vision of the future of visual quality assessment research, with a small number of IQA measures reviewed. Chandler [5] gave a systematic review in 2013. The first half of this review discusses key properties of visual perception, IQA databases, existing IQA algorithms. The second half of this article highlights several open challenges in the field. Generally, the survey focuses more on the problems and challenges of IQA rather than review of algorithms. In 2013, He et al. [6] briefly introduced the history and developments of IQA metrics, with a relatively small number (about two dozens) of IQA metrics involved.

In 2014, Mohammadi et al. [7] reviewed both subjective and objective IQA methods, with a focus on 9 classic FR IQA measures, as well as two emerging directions: high dynamic range (HDR) and 3D IQA. Manap and Shao [8] reviewed generally-purpose NR IQA in 2015, also with small number of very popular algorithms. In 2017, Xu et al. [9] reviewed both distortion-specific and general-purpose NR IQA measures, but the scope of the reviewed measures was also limited.

Besides the objective IQA measures, some review studies have summarized the subjective IQA databases. Winkler [10] gave an analysis of public image and video databases for visual quality assessment in 2012, with a comparison of source contents, distortion processes, and subjective ratings. In 2013, Winkler and Subramanian [11] gave an overview of eye tracking datasets, with some of which related to IQA because IQA and visual attention are two closely related research areas.

1.2 Need for a new survey

A comprehensive and up-to-date survey is needed because the current surveys have the following limitations.

First, the reviewed IQA algorithms are small in number and restricted in scope. Most surveys focus on the classical and well-known IQA measures. Classical IQA measures are important, but a comprehensive overview of IQA measures is also indispensable. Chandler [5] cited more than 300 papers, but this survey focused more on the related areas, problems, and challenges of IQA rather than the algorithms. Some existing surveys concentrate on specific areas, e.g., NR IQA [8, 9].

Second, the reviewed IQA measures are not up-to-date. The most recent 2 surveys were published in 2015 and 2017 [8, 9]. But these two surveys focused only on NR IQA. Other surveys were published more than 3 years ago. The recent years have witnessed a great development of IQA. Many new objective IQA measures have been proposed, which should be analyzed and compared.

Last, some new emerging areas in IQA research are not included. For example, stereoscopic IQA, saliency-guided IQA, screen content IQA, tone-mapping IQA, multi-exposure fusion IQA, retargeting IQA, multiply distorted IQA, authentic distortion IQA, dehazing IQA, virtual reality IQA, and many other topics, all have been intensively researched in recent years. These emerging topics represent new trends of IQA, but they are missing from almost all current IQA surveys.

Considering the limitations of the current IQA surveys described above, a comprehensive and up-to-date survey, including not only classical IQA measures but also emerging topics, is in great need for a better top-down understanding of the history, current state-of-the-art, and future trend of the field. This survey is written to fulfill that need.

1.3 Scope and organization of this survey

As limited by pages, we confine our survey pool to encompass only papers published in the last 16 years, and we specifically include these papers that are considered important but have not been included in existing surveys. In Section 2, the subjective IQA databases, which are sub-categorized into traditional and emerging ones, are reviewed. Following that, the traditional and emerging objective IQA measures are reviewed in Sections 3 and 4, respectively. Emerging IQA measures are classified according to specific applications, including stereoscopic IQA, saliency-guided IQA, screen content IQA, tone-mapping IQA, multi-exposure fusion IQA, retargeting IQA, multiply distorted IQA, and authentic distortion IQA, dehazing IQA, virtual reality IQA, and many other emerging topics. In Section 5, evaluation process of IQA measures is discussed, with a comparison of their performances. Section 6 concludes this paper. A framework of the included topics is illustrated in Figure 1 for an easy navigation through the paper.

2 Subjective image quality assessment databases

Subjective assessment is the most reliable way to evaluate the quality of images, because human eyes are usually the ultimate receiver of the images. Subjective quality assessment involves the processes of viewing environment setup, subjects recruitment, subjects grading, and processing of the subjective results, as suggested by ITU-R BT.500 [12]. The subjective IQA databases are used to train and test objective IQA metrics. In this section, a total of 40 IQA databases are reviewed, which are categorized into traditional and emerging databases according to the image content types and the underlying applications. Basically, traditional IQA databases are built for general-purpose IQA, whereas emerging IQA databases are generally designed for some specific IQA problems and applications.

Table 1 [13–41, 41–51] gives an overview of 40 databases that are widely used in the research of visual quality assessment. Information including the numbers of reference images, distorted images, distortion types and image resolutions, as well as subjective score types is summarized. Traditionally general databases generally include regular images distorted by some common distortion types, e.g., JPEG and JPEG2000 compression, white noise injection, and blurring. While the emerging IQA databases include 3D image, retargeting image, multiple distorted image, screen content image, authentic distortion image, tone-mapping, view synthesis, dehazing, virtual reality image databases. Detailed information of these databases can be found below, including the number of source images and distorted images, the distortion types as well as specifications of subjective viewing tests.

2.1 Traditional databases

- LIVE image quality assessment database [13]. LIVE includes 29 pristine images and 779 distorted images corrupted by 5 types of distortions, i.e., JPEG compression (JPEG), JPEG2000 compression (JP2K), white noise (WN), Gaussian blur (GB), and simulated fast fading Rayleigh channel (FF). Each distortion type contains 5 or 4 distortion levels. Most images are 768×512 pixels in size.
- Tampere image database 2008 (TID2008) [14]. TID2008 includes 25 pristine images and 1700 distorted images corrupted by 17 types of distortions, with 4 levels for each distortion type. All images have a fixed resolution of 512×384.
- Tampere image database 2013 (TID2013) [15]. TID2013 is extended from TID2008 by increasing the number of distortion levels to 5, and the number of distortion types to 24. Therefore, 3000 distorted images are generated from 25 pristine images. The subjective testing and data processing steps are similar to that of TID2008.
- Categorical subjective image quality (CSIQ) database [16]. It contains 30 pristine images and 866 distorted images corrupted by JPEG, JP2K, WN, GB, additive pink Gaussian noise, and global contrast decrements, with 5 or 4 levels for each distortion type. The resolution is 512×512.
- IRCCyN/IVC subjective quality assessment database [17]. IVC consists of 10 pristine images and 235 distorted images corrupted by JPEG, JP2K, blur, and locally adaptive resolution coding, with 5 levels

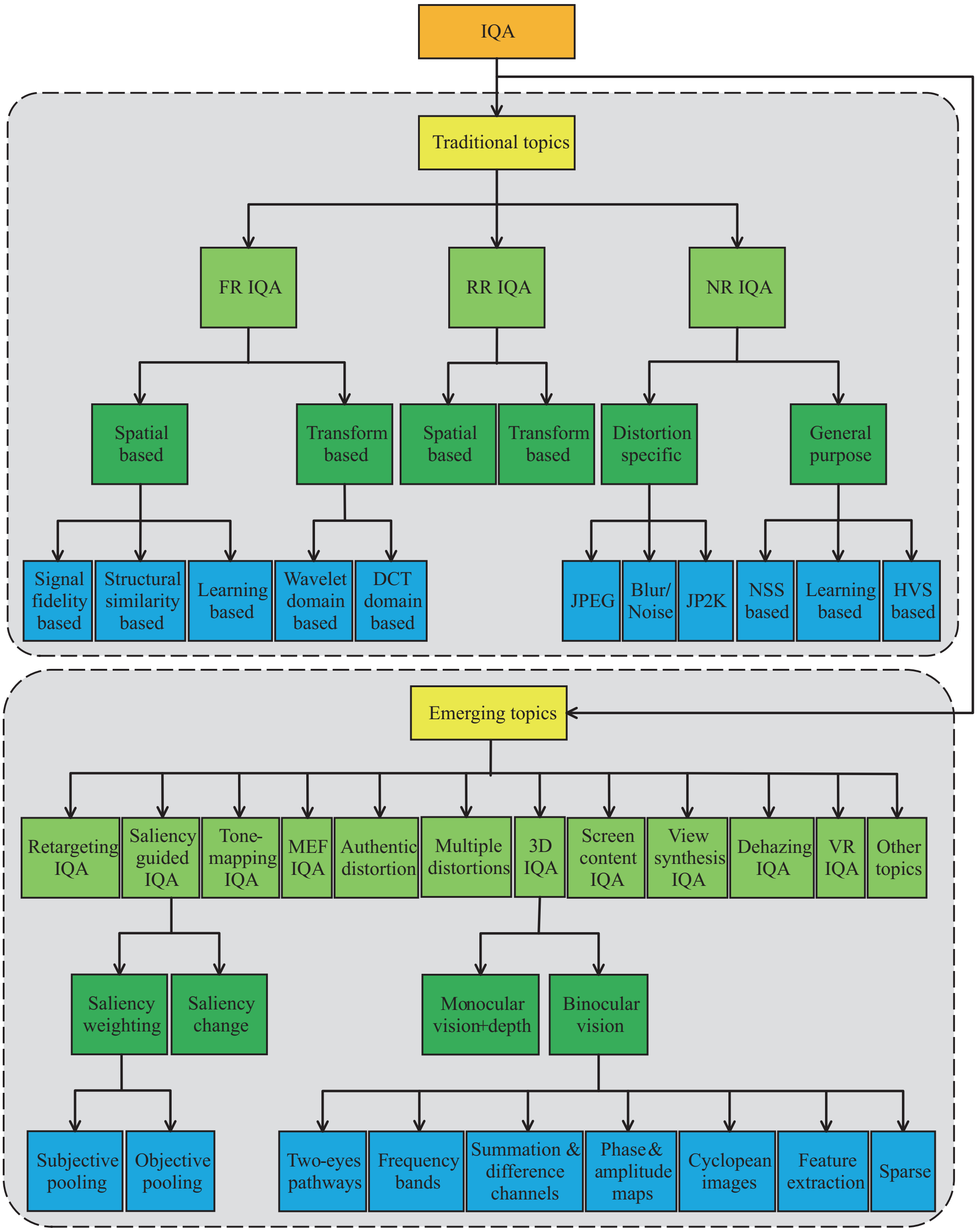


Figure 1 (Color online) Scope of this paper.

for each distortion type. The resolution is fixed at 512×512.

• MICT image quality evaluation database [18]. MICT includes 14 pristine images and 168 distorted images corrupted by JPEG and JP2K, with 6 levels for each distortion type. The resolution is 768×512.

• A57 database [19]. A57 includes 3 pristine images and 54 distorted images corrupted by 6 types of distortions, with 3 levels for each distortion type. All images are in gray scale. The resolution is 512×512.

• Waterloo exploration database (WED) [20]. WED includes 4744 pristine natural images and 94880 distorted images corrupted by JPEG, JP2K, GB, and WN, with 5 levels for each distortion type. The images have various resolutions. No human opinion score is provided, but the authors introduce several alternative test criteria to evaluate the IQA models.

Table 1 An overview of IQA databases[a)]

Category	Database	Year	No. Ref.	No. Dist.	No. Dist. Type	No. Dist. Level	Resolution	Ground-truth
General	LIVE [13]	2004	30	779	5	5 or 4	$\sim 768 \times 512$	DMOS
	TID2008 [14]	2008	25	1700	17	4	512×384	MOS
	TID2013 [15]	2013	25	3000	24	5	512×384	MOS
	CSIQ [16]	2009	30	866	6	5 or 4	512×512	DMOS
	IVC [17]	2005	10	54	4	5	512×512	MOS
	MICT [18]	2001	14	168	2	6	768×512	MOS
	A57 [19]	2007	3	54	6	3	512×512	MOS
	WED [20]	2017	4744	94880	4	5	–	–
3D	IVC 3D [21]	2009	6	90	3	5	512×448	DMOS
	LIVE 3D Phase I [22]	2013	20	365	5	–	640×360	DMOS
	LIVE 3D Phase II [23]	2013	8	360	5	–	640×360	DMOS
	Waterloo 3D Phase I [24]	2015	6	330	3	–	$\sim 1300 \times 1100$	MOS
	Waterloo 3D Phase II [24]	2015	10	460	3	–	1920×1080	MOS
	Ningbo 3D Phase I [25]	2009	10	400	4	–	$\sim 1300 \times 1100$	DMOS
	Ningbo 3D Phase II [26]	2011	12	312	5	–	480×270 to 1024×768	DMOS
	Tianjin 3D [27]	2009	30	270	3	–	320×240 to 1024×768	DMOS
	MCL-3D Database [28]	2015	9	693	7	4	1024×728 or 1920×1080	MOS
	IVY 3D [29]	2013	–	120	1	–	1920×1080	MOS
	MMSP 3D [30]	2010	–	54	1	6	1920×1080	MOS
Retargeting	MIT RetargetMe [31]	2012	37	296	8	–	–	Pair-wise
	CUHK Retargeting [32]	2012	57	171	10	–	–	MOS
Multiple Distortions	LIVEMD [33]	2012	15	405	2	–	1280×720	DMOS
	MDID2013 [34]	2013	12	324	–	–	768×512 or 1280×720	DMOS
	MDID2016 [35]	2016	20	1600	–	–	512×384	MOS
Screen content	SIQAD [36]	2014	20	980	7	7	$\sim 700 \times 700$	DMOS
	SCIQ [37]	2017	40	1800	9	5	1280×720	MOS
	CCT [38]	2017	72	1320	2	11	1280×720 to 1920×1080	MOS
	HSNID [39]	2019	20	600	6	5	–	MOS
Authentic	LIVE Wild [40]	2016	0	1162	–	–	500×500	MOS
	CID2013 [41]	2015	0	480	–	–	1600×1200	MOS
Tone-mapping & MEF	TMID [42]	2013	15	120	8	–	–	Rank
	ESPL-LIVE HDR [43]	2017	605	1811	11	–	960×540 to 304×540	MOS
	MEF [44]	2015	17	136	8	–	340×512	MOS
View synthesis	IRCCyN/IVC DIBR [45]	2011	12	84	7	–	1024×768	DMOS
Dehazing	DHQ [46]	2019	250	1750	7	–	–	MOS
	SHRQ [47]	2019	75	600	8	–	–	MOS
	IVC Dehazed Image [48]	2015	25	200	8	–	–	MOS
VR	OIQA [49]	2018	16	320	4	5	11332×5666 to 13320×6660	MOS
	CVIQ [50]	2019	16	528	3	11	4096×2048	MOS
	LIVE 3D VR IQA [51]	2019	15	450	6	5	4096×2048	DMOS

a) No.: Number of; Ref.: Reference; Dist.: Distortion.

2.2 Emerging databases

The databases for emerging IQA problems in this subsection are organized by applications in accordance with the algorithms to be reviewed in Section 4.

2.2.1 *3D IQA databases*

• IRCCyN/IVC 3D image quality database [21]. IVC 3D is the first public 3D image quality database. A total of 6 stereo images and 90 degraded images are corrupted by symmetrical JPEG, JP2K, and GB distortions, with 5 levels for each distortion type. The resolution is 512×448.

• LIVE 3D image quality database [22, 23]. LIVE 3D is introduced in two phases. Phase I includes 20 pristine stereopairs and 365 symmetrically distorted stereopairs, while phase II includes 8 pristine stereopairs and 360 symmetrically and asymmetrically distorted stereopairs. Both phases use the five distortion types similar to the LIVE IQA database. The resolution is 640×360.

• Waterloo IVC 3D image quality database [24]. Waterloo IVC 3D is also introduced in two phases. Phase I includes 6 pristine stereopairs and 78 distorted single-view images and 330 symmetrically and asymmetrically distorted stereopairs. The resolution is 1390×1100 or 1342×1100. Phase II includes 10 pristine stereopairs and 130 distorted single-view images and 460 symmetrically and asymmetrically distorted stereopairs. The resolution is 1920×1080. In both phases, WN, GB, and JPEG are introduced as distortions.

• Ningbo University database [25, 26, 52]. NBU database is also introduced in two phases. Phase I includes 10 pristine stereopairs 400 asymmetrically distorted stereopairs corrupted by JPEG, JP2K, WN, and GB. The resolution is around 1300×1100. Phase II includes 12 pristine stereopairs 312 symmetrically distorted stereopairs corrupted by JPEG, JP2K, WN, GB, and H.264 compression. The image resolution varies from 480×270 to 1024×768.

• Tianjing University database [27]. TJU database contains a total of 30 pristine stereopairs and 270 distorted stereopairs corrupted by symmetrical JPEG, JP2K, and WN. The image resolution varies from 320×240 to 1024×768.

• MCL-3D database [28]. MCL-3D has 693 stereoscopic image pairs generated from 9 image-plus-depth sources. Distortions including JPEG, JP2K, GB, WN, down-sampling blur, and transmission errors are added to either the texture image or the depth image. Besides the above distortions, imperfect rendering is also included. Four levels of distortions are added for each type. The resolution is either 1024×728 or 1920×1080.

• IVY LAB 3D image database [29]. In IVY LAB 3D, there are 120 real scenes and human-labeled visual discomfort provided in a form of MOS. The magnitude of maximum crossed disparity ranges from $0.11°$ to $5.07°$, corresponding to 0 to 285 pixels. The resolution is 1920×1080.

• MMSP 3D image quality assessment database [30]. MMSP 3D has 9 stereoscopic scenes, and each scene is captured with 6 different inter-camera distances. All images are JPEG-compressed with a resolution of 1920×1080.

2.2.2 *Retargeting IQA databases*

• MIT RetargetMe database [31]. RetargetMe contains 37 source images which have one or more of the six major retargeting attributes: line/edge, face/people, foreground objects, texture, geometric structures and symmetry. The retargeted images are created using 8 retargeting operators with 25% or 50% of width or height reduction. Paired comparison tests are conducted, and the corresponding numbers of times that the retargeted image is favored over others are provided as the subjective scores.

• CUHK image retargeting subjective database [32]. The CUHK database includes 171 retargeted images generated from 57 natural images. Two more retargeting operators than RetargetMe are used. Each image is retargeted with 25% or 50% width/height reduction.

2.2.3 *Multiply distorted IQA databases*

• LIVE multiply distorted (LIVEMD) database [33]. LIVEMD consists of 15 reference images and 405 multiply distorted images. The database includes one/double-fold artifacts. Each multiply distorted images are corrupted under two multiple distortion scenarios: GB followed by JPEG and GB followed by WN. All images have a resolution of 1280×720.

• Multiply distorted image database 2013 (MDID2013) [34]. MDID2013 has a total of 12 pristine images and 324 distorted images. Each pristine image is corrupted successively by GB, WN, and JPEG. The images have resolutions of 768×512 or 1280×720.

• Multiply distorted image database 2016 (MDID2016) [35]. MDID2016 consists of 20 reference images and 1600 distorted images. Five distortion types are introduced, i.e., WN, GB, JPEG, JP2K, and contrast change (CC). The order of distortions is as follows: GB or CC first, JPEG or JP2K second and WN last. All distorted images are with random types and levels of distortions. The image resolution is 512×384.

2.2.4 *Screen content IQA databases*

- Screen image quality assessment database (SIQAD) [36]. SIQAD includes 20 pristine and 980 distorted screen content images (SCIs). Distortion types include WN, GB, CC, JPEG, JP2K, motion blue (MB), and layer segmentation based compression, with 7 levels for each type. The images have various resolutions near 700×700.
- Screen content image quality (SCIQ) database [37]. SCIQ consists of 40 pristine and 1800 distorted SCIs corrupted by 9 types of distortions, including WN, GB, MB, CC, JPEG, JP2K, color saturation change (CSC), color quantization with dithering (CQD), and the screen content coding extension of high efficiency video coding (HEVC-SCC). Five distortion levels are considered. The resolution is fixed at 1280×720.
- Cross-content-type (CCT) database [38]. CCT is constructed to conduct cross-content-type IQA research. CCT consists of 72 pristine and 1320 distorted natural scene images (NSIs), computer graphic images (CGIs), and SCIs. Two distortion types are considered, i.e., HEVC and HEVC-SCC coding, with 11 distortion levels for each type. The image resolution is either 1920×1080 or 1280×720.
- Hybrid screen content and natural scene image database (HSNID) [39]. HSNID has 10 pristine NSIs and 10 pristine SCIs, and 600 distorted NSIs and SCIs corrupted by WN, GB, MB, CC, JPEG, and JP2K, with 5 distortion levels for each type.

2.2.5 *Authentic distortion IQA databases*

- LIVE in the wild image quality challenge database [40]. It includes 1162 authentically distorted images captured using a variety of mobile devices. Complex real distortions, which are not well-modeled by the synthetic distortions are included. All images are cropped to the resolution of 500×500. MOSs collected via crowdsourcing are provided.
- Camera image database (CID2013) [41]. CID2013 is designed to test no-reference IQA algorithms. It includes 480 real images captured from 8 typical scenes using 79 consumer cameras and mobile phones. The images are rated from 5 aspects: the overall quality, sharpness, graininess, lightness, and color saturation scales. The images are scaled to a size of 1600×1200.

2.2.6 *Tone-mapping IQA databases*

- Tone-mapped image database (TMID) [42]. TMID is composed of 120 tone-mapped images generated from 15 sets of HDR images using 8 different tone mapping operators (TMOs). Subjects are asked to rank the 8 images in each image set, and the mean ranking scores within the sets are provided.
- ESPL-LIVE HDR image database [43]. It consists of 1811 HDR-processed images created from 605 high quality source HDR scenes using tone-mapping and multi-exposure fusion (MEF). Post-processing artifacts of HDR image creation are also considered. A total of 11 processing techniques are used. The image resolution is 960×540 or 304×540. MOSs collected via crowdsourcing are provided.

2.2.7 *Multi-exposure fusion IQA databases*

- Multi-exposure fusion image database [44]. The database includes 136 fused images created from 17 image sets using 8 MEF algorithms. The images have resolutions near 340×512. Eight fused images corresponding to the same image set are showed to the subjects at the same time.
- ESPL-LIVE HDR image database [43]. The database includes 710 images created via MEF.

2.2.8 *View synthesis IQA database*

- IRCCyN/IVC DIBR image database [45]. Depth-image-based-rendering (DIBR) tries to synthesize a view for any viewpoint given the sparse views and the scene depth. The database consists of 12 reference images and 84 synthesized ones created by 7 different DIBR algorithms. All images share the resolution of 1024×768, and DMOSs are provided as quality scores.

2.2.9 *Dehazing IQA databases*

- Dehazing quality (DHQ) database [46]. The DHQ database includes 1750 dehazed images generated from 250 real hazy images of various haze densities using 7 representative image dehazing algorithms. All dehazed images are labeled with human rated MOSs.
- Synthetic haze removing quality (SHRQ) database [47]. It consists of two subsets: regular and aerial image subsets, which include 360 and 240 dehazed images created from 45 and 30 synthetic hazy images using 8 image dehazing algorithms, respectively. Original haze-free images, dehazed images, and the MOSs of each dehazed images are provided for dehazing IQA studies.
- Waterloo IVC dehazed image database [48]. A total of 225 images are provided in this database, including 25 hazy images and 200 dehazed images created from these 25 hazy images using 8 dehazing algorithms. MOSs of all 225 images are provided.

2.2.10 *Virtual reality IQA databases*

- Omnidirectional IQA (OIQA) database [49]. The OIQA database includes 16 source omnidirectional images and 320 distorted ones degraded by 4 common distortion types, namely JPEG, JP2K, GB, and WN. For each distortion type, 5 levels of distortions are introduced. MOS is provided for each distorted image. Besides the images and MOSs, the head and eye movement data are also provided in the database. The resolutions of images range from 11332×5666 to 13320×6660.
- Compressed VR image quality (CVIQ) database [50, 53]. It consists of 16 reference VR images and 528 compressed VR images generated by using 3 types of compression, including JPEG, H.264, and H.265. For each distortion type, 11 levels of distortions are introduced. All images are provided at the resolution of 4096×2048. MOSs of all distorted VR images are also provided.
- LIVE 3D VR IQA database [51]. A total of 450 distorted images obtained from 15 reference 3D virtual reality (VR) images degraded by 6 types of distortions, including WN, GB, downsampling distortion, stitching distortion, VP9 compression, and H.265 compression. Most images are provided at the resolution of 4096×2048. DMOSs of all distorted images are provided.

2.2.11 *Visual attention databases for IQA*

- TUD image quality database: eye-tracking release 1 [54]. This database provides eye-tracking data of the 29 pristine images from the LIVE under free-viewing conditions.
- TUD image quality database: eye-tracking release 2 [55]. The database is constructed to study how people look at images when assessing image quality. Eye-tracking data is collected under both free-viewing and quality rating conditions.
- TUD image quality database: interactions [56]. The database is constructed to investigate the deviations of quality scoring saliency from free looking saliency.
- Visual attention for image quality (VAIQ) database [57]. VAIQ provides visual attention data of the pristine images from the LIVE [13], MICT [18], and IVC [17] databases.
- Visual attention data for IQA databases [58]. It provides visual attention data of the pristine images from seven widely used IQA databases including the LIVE [13], TID2008 [14], CSIQ [16], MICT [18], LIVE MD [33], IVC [17], and A57 [19] databases.

3 Objective image quality assessment: traditional topics

In this section, we review the traditional objective IQA measures. These measures are built for general-purpose IQA, and they are assumed to be able to handle various kinds of distortions, for example the

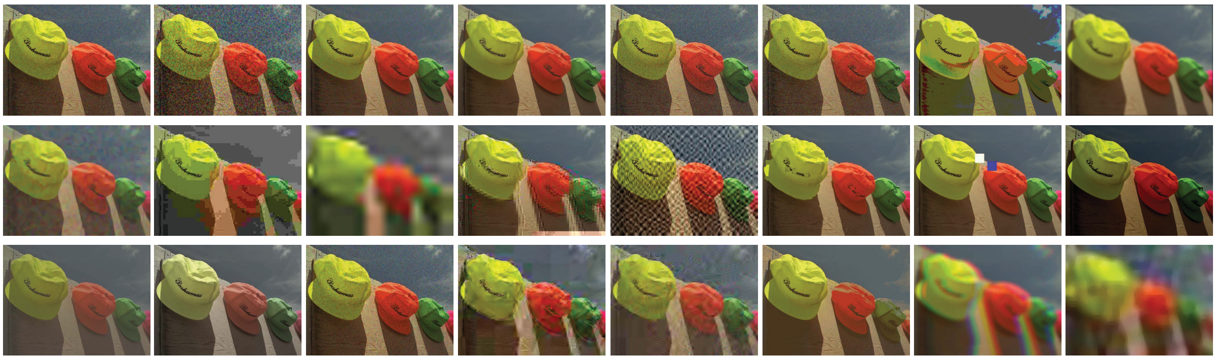


Figure 2 (Color online) An illustration of various general distortion types. All distortions included in the TID2013 database [15] are shown in this figure.

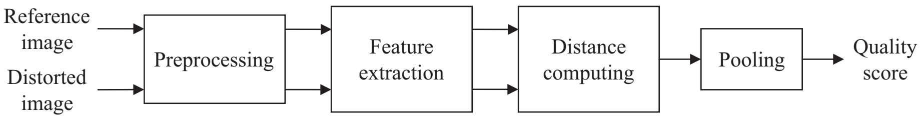


Figure 3 General framework of FR IQA algorithms. Features are extracted from both images, and then the feature distance is calculated.

distortions shown in Figure 2. According to the availability of the distortion-free reference image, IQA measures can be categorized as FR, RR, and NR [1,2,4–9]. Each category will be reviewed in this section.

3.1 Full-reference IQA algorithms

The objective of FR IQA is to predict the quality of a target image with full access to the original reference image. For the most straightforward type of IQA task, many FR IQA measures follow a similar framework, i.e., feature extraction from both images followed by distance calculation as illustrated in Figure 3. The features can be collected from either spatial or transform domain, or both. Because feature extraction is the key step for FR IQA measures, we review IQA metrics using the underlying features as a lead in the following subsections.

3.1.1 *Spatial domain methods*

(1) Signal fidelity. Although traditional signal fidelity measures like mean square error (MSE) and peak signal-to-noise ratio (PSNR) are often challenged because they have no consideration of characteristics of image signal and the HVS, they are still widely used as FR measures [1]. Given a reference image R and its distorted version D, MSE is defined as

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[R(i,j) - D(i,j)]^2, \tag{1}$$

where i, j are pixel indexes, and M, N are image height and width. Then PSNR can be written as

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{MAX_I}^2}{\mathrm{MSE}}\right), \tag{2}$$

where $\mathrm{MAX_I}$ is the maximum possible pixel value of the image.

(2) Structural similarity and variants. Considering that the HVS is highly sensitive to the structural information in images, some measures extract image structures and then compute structural similarity as the quality. Wang et al. [59] proposed the structural similarity (SSIM) measure, whose framework is illustrated in Figure 4 [59]. In SSIM, the luminance, contrast and structure features are extracted, and then the SSIM index is calculated as from the reference image and the distorted image,

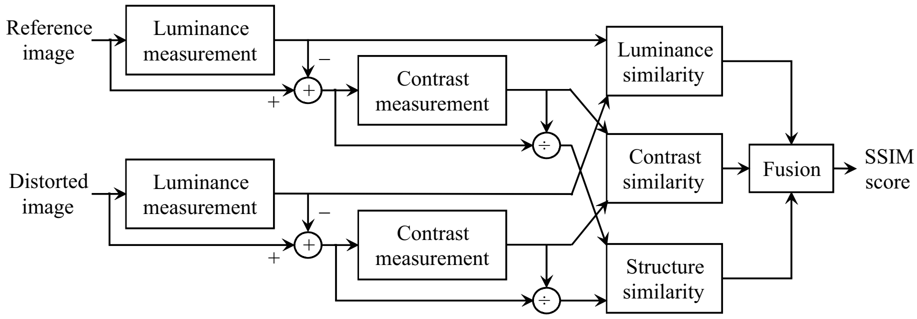


Figure 4 Framework of the FR SSIM index [59], which measures the image luminance, contrast and structure similarity.

and then the obtained features are integrated by a pooling strategy to derive the quality score:

$$\begin{aligned}\mathrm{SSIM}(x,y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2+\mu_y^2+c_1}\cdot\frac{2\sigma_x\sigma_y+c_2}{\sigma_x^2+\sigma_y^2+c_2}\cdot\frac{\sigma_{xy}+\frac{c_2}{2}}{\sigma_x\sigma_y+\frac{c_2}{2}}\\ &= \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)},\end{aligned} \tag{3}$$

where the first, second, and third terms in the first line of (3) measure image luminance, contrast and structure similarity, respectively; μ, σ, σ_{xy} are local mean, variance, and covariance; x, y indicate two images; c_1, c_2 are small two stabilization constants.

Because the HVS is known to have a multi-resolution processing paradigm for visual inputs, SSIM is extended to multi-scale SSIM (MS-SSIM) [60]. In SSIM and MS-SSIM, the distortion or feature similarity map is pooled using an averaging pooling. Wang and Li [61] proposed an information content weighting strategy for pooling, and used it to improve MSE, PSNR and SSIM. Tan et al. [62] proposed a perceptually meaningful MSE-SSIM measure by analyzing the relationship between the MSE and SSIM under an additive noise distortion model. MSE-SSIM is expressed in terms of the variance of the reference image and the MSE between the reference and distorted images. Wu et al. [63] used the internal generative mechanism (IGM) of human brain in pooling: the images are decomposed into the predicted portion and disorderly portion and then SSIM and PSNR are applied to the two portions respectively.

The HVS largely relies on edge information for image interpretation, therefore, image gradients are found to be very effective for perceptual quality metric. The structural similarity can also be computed from the gradient domain. Zhang et al. [64] proposed a feature similarity (FSIM) index. In FSIM, phase congruency (PC) and gradient magnitude (GM) for both images are first extracted, and the similarity is calculated using the basic form:

$$s(x,y) = \frac{2f_xf_y + c}{f_x^2+f_y^2+c}, \tag{4}$$

where f indicates the feature (PC or GM), c is a stabilization constant. PC similarity and GM similarity are then multiplied to the FSIM. A perceptual similarity (PSIM) measure is proposed in [65] by measuring the similarities of micro- and macro-structures which are described by GM maps. Similar to FSIM, color similarity is also incorporated to strength the PSIM measure. Liu et al. [66] proposed a gradient similarity (GSIM) index by comparing the similarity of the gradient values using the similarity form shown in (4). Xue et al. [67] first computed gradient magnitude similarity (GMS) map, and then designed a gradient magnitude similarity mean (GMSM) measure via average pooling and a gradient magnitude similarity deviation (GMSD) by computing the standard deviation of the GMS map. Zhu et al. [68] proposed a measure using visual gradient similarity (VGS). They proposed a multi-scale global contrast registration method, and conducted a point-wise comparison by multiplying the similarity of gradient direction and magnitude of reference and distorted images.

(3) Other structure and edge based features. Zhan et al. [69] evaluated image quality through combining the distribution of different structural distortion types and the degree of structural differences. Zhang et al. [70] proposed a non-shift edge based ratio (NSER) method. The authors utilized the variation of the number of edge points in non-shift edge map to measure the quality. Capodiferro et al. [71] proposed a method by integrating a structure loss measure and a categorical indicator of impairment types. Fisher information theory is used to calculate the positional structural information to estimate the quality, and the distortion categorical index is estimated by the mutual correlation of gradients of two images. Di Claudio et al. [72] utilized the combination of two separate metrics to measure the perceptual impact of detail losses and of spurious details. Ding et al. [73] measured the dominant structural information change using anisotropy and local directionality. Sun et al. [74] evaluated image quality based on superpixel luminance similarity, superpixel chrominance similarity, and pixel gradient similarity.

(4) Learning-based feature extraction and integration. Instead of directly combining quality related image features, some IQA metrics are based on learning techniques for feature discovery and integration. An obvious advantage of using machine learning technique in feature integration is that the model can be mathematically optimal and therefore has superior performance.

Narwaria et al. [75] used singular value decomposition (SVD) based features, and the support vector regression (SVR) was used as a feature fusion tool. Another SVD-based measure was proposed in [76], and the authors first calculated the distance between the singular values of the reference image blocks and distorted image blocks, then they computed a global value from each block to represent the final quality. Liu et al. [77] introduced a novel parallel boosting measure which inherited the advantages of some state-of-the-art FR measures. Specifically, the authors utilized the SVR to integrate the quality features extracted by state-of-the-art FR measures. Wang et al. [78] utilized the parts-based representation of non-negative matrix factorization (NMF) to estimate image distortions, then the extreme learning machine (ELM) was used to generate the final quality score. Peng et al. [79] introduced a two-stage framework based on support-vector classification and k-nearest-neighbor regression. Specifically, the authors devised a probabilistic approach to make use of distortion-specific features, and the authors conducted a decision fusion to integrate the SSIM, visual signal-to-noise ratio (VSNR) and visual information fidelity (VIF) measures by k-nearest-neighbor regression.

Sparse representation is another widely used learning related method for feature selection. Chang et al. [80] proposed a sparse feature fidelity (SFF) measure, with sparse representation and independent component analysis (ICA) based feature selection followed by similarity measurement. Li et al. [81] proposed an image quality index with adaptive sub-dictionaries (QASD) based on sparse representation. The authors trained an over-complete dictionary from natural images and adaptively selected the sub-dictionaries to extract sparse features. The similarity of two images' sparse features and some other auxiliary features is computed as the quality index. Yuan et al. [82] utilized sparse representation in local image patches, and the difference between the sparse representation of reference and distorted patches was computed as the quality map. The kernel ridge regression (KRR) was used to fuse the local quality into the final quality score. Ahar et al. [83] first deployed a Fourier basis for sparse coding, then ranked the amplitudes of the sparse coefficients, and finally assessed the correspondence between ranked coefficients of the reference and the distorted images.

Besides those basic learning techniques, some methods utilize more complicated deep learning to predict the quality, considering the successes of deep learning in various visual problems [84–89]. Gao et al. [90] proposed a deep similarity index (DeepSim) with deep neural network (DNN). The authors estimated the local similarities of DNN features between two images, and pooled the local similarities to the final quality. Wang et al. [91] first utilized a local linear model (LLM) to detect the degradation between reference and distorted images, and then introduced a distortion-specific compensation strategy to deal with the offset caused by different image distortion types. In their method, convolution neural network (CNN) was used to compute the score offset. Kim et al. [92] proposed a CNN based FR IQA model, in which the optimal visual weights were learned based on the understanding of database information itself. Bosse et al. [93] introduced a neural network-based approach to FR and NR IQA which allows for feature learning and regression in an end-to-end manner. It learned the local quality and local weights jointly, and it can be used for FR or NR IQA with slight adaptations.

3.1.2 *Transform domain methods*

Some methods perform transformation to extract features, because the quality degradation can be also reflected in the transformation domain.

(1) Wavelet domain. Sheikh et al. [94] explored IQA from the perspective of information processing of the HVS. Specifically, they adopted two different natural scene statistics (NSS) models to describe the reference and distorted images. The reference image is expressed by Gaussian scale mixture (GSM) in the wavelet domain, while the distorted image is expressed by a simple signal attenuation and additive Gaussian noise model in the wavelet domain. The final information fidelity criterion (IFC) is derived from the mutual information between the source and distorted images. Sheikh et al. [95] later extended their study and investigated the relationship between image information and perceptual quality. They formulated the quality perception process with a distortion channel followed by a HVS channel and proposed a VIF measure which quantifies the information error between reference and distorted images. VIF is derived from the quantification of two types of mutual information: the mutual information between the input and the output of the HVS channel (described via a stationary white Gaussian noise model) when no distortion channel is presented (i.e., reference mutual information), and the mutual information between the input of the distortion channel and the output of the HVS channel for the test image. The NSS models used to describe the reference and distorted images are similar to IFC [94].

Demirtas et al. [96] developed a multi-scale quality estimator for images with different spatial resolutions to tackle the problem that images can be viewed at different distances on different devices. Image luminance is first decomposed into subbands by wavelet transformation, then the wavelet subbands are described via Gaussian scale mixture, and mutual information between reference and distorted images' subbands is estimated as the quality.

Chandler et al. [19] proposed a VSNR measure. Wavelet transform is used to determine detectability of the degradation. Low-level property of perceived contrast and the mid-level property of global precedence are considered for supra-threshold distortions. VSNR is computed via a simple sum of the Euclidean distances in distortion-contrast space of wavelet coefficients.

Li et al. [97] classified the differences between the reference and distorted images as detail loss and additive impairment, which refer to the loss of useful visual information and the redundant visual information, respectively. The authors adopted a wavelet-domain decoupling algorithm to separate the detail loss and additive impairment features before pooling.

Tang et al. [98] proposed an algorithm based on α-stable model similarity in wavelet domain. The authors found that the leptokurtic and heavy-tailed behaviors of image wavelet coefficients are characterized by symmetric α-stable density, and the model parameters are highly correlated to the distortions. Thus they normalized the characteristic function of symmetric α-stable model to derive the final similarity measurement.

(2) DCT domain. Bae and Kim [99] proposed a DCT based quality degradation metric (DCT-QM), which was derived from a psychophysics theory for low-level mechanism of neural receptive responses in visual cortex. It was computed as a weighted mean ℓ_2 norm in the DCT domain, which was easy to implement and had some desirable mathematical properties, including differentiability, convexity, and valid distance metrics. In a related study [100], Bae and Kim proposed a structural contrast-quality index (SC-QI) based on a structural contrast index which can describe the local and global visual quality perception of the image with various distortion types. The feature values of the structural contrast index are calculated from the DCT coefficients of the image luminance. The authors compared the similarity of contrast energy values in low frequency, middle frequency and high frequency regions in 4×4 DCT blocks. They also modified SC-QI to a structural contrast distortion metric (SC-DM) to retain some favorable mathematical properties.

3.2 Reduced-reference IQA algorithms

The objective of RR IQA measures is to predict the quality of image with limited access to the reference image. Generally RR features of the reference are extracted at the sender side and transmitted to the

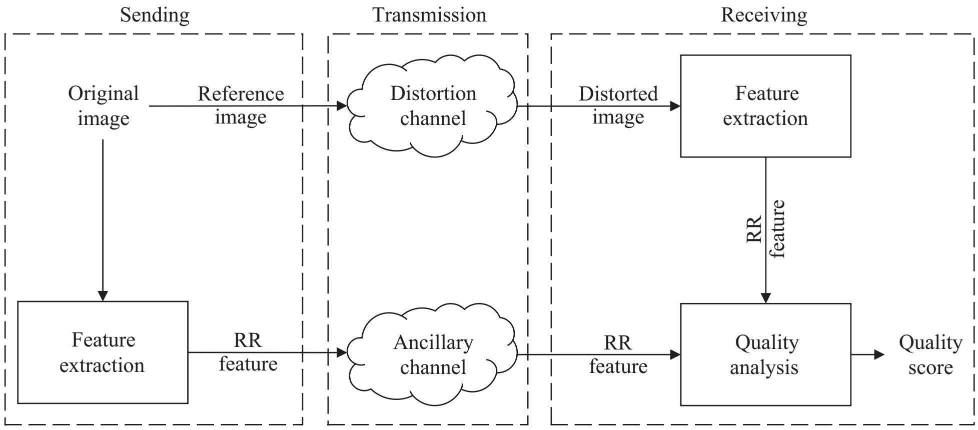


Figure 5 General framework of RR IQA algorithms. RR features of the reference and distorted images are extracted at the sender and receiver side, respectively. Then the RR features of the reference and distorted images are used collectively to compute the quality.

receiver side. The same feature extraction process is performed for the distorted image at the receiver side, and the RR features of the reference and distorted images are used collectively to compute the quality of the distorted image, as illustrated in Figure 5. A good RR approach should be able to obtain high quality prediction accuracy with the limited amount of RR features. We also classify RR metrics into spatial domain based and transfer domain based methods.

3.2.1 *Spatial domain methods*

Redi et al. [101] extracted color correlogram features which described spatial correlations of colors. The color distribution features are mapped into a numerical expression which describes the perceived quality. Wu et al. [102] decomposed images into orderly and disorderly portions. Then the quantities of visual information of the two portions are computed. Finally, information fidelities on the two portions are evaluated and combined into the overall quality score. Decherchi et al. [103] introduced an augmented version of the basic extreme learning machine (ELM), the circular-ELM (C-ELM). The second order statistics of color information is employed as features to describe the image. The final prediction of the quality score is conducted by the trained C-ELM. Bampis et al. [104] used Gaussian scale mixture models to describe images which have been locally mean subtracted. The locally weighted entropies of distorted images and reference images are calculated. The differences between entropys are averaged over blocks, thereby yielding the image quality score. Zhang et al. [105] proposed an RR IQA method using local sharpness because multi-scale local sharpness maps are affected differently by different distortion types. Min et al. [106] introduced a saliency-induced RR IQA method because different types and levels of degradation can strongly influence saliency detection. Liu et al. [107] proposed an RR IQA method based on free-energy principle and sparse representation.

3.2.2 *Transform domain methods*

Gao et al. [108] used the wavelet-based contourlet transform to process the images. Normalization is conducted in the transform domain considering the contrast sensitivity function, then the difference of proportions of visual sensitivity coefficients in distorted and reference images is calculated as the quality score. Wang et al. [109] utilized generalized Gaussian density (GGD) function to fit the marginal distribution of coefficients computed from the wavelet subbands derived by steerable pyramid decomposition, and model parameters are used as features to predict the quality score. Soundararajan et al. [110] found that the wavelet coefficients of reference and distorted images followed the Gaussian scale mixture (GSM) distribution. Parameters of GSM are estimated to measure the differences of entropys between distorted and reference images. Rehman et al. [111] extracted statistical features from images after the multiscale multiorientation divisive normalization transformed and measured the quality as the distance between

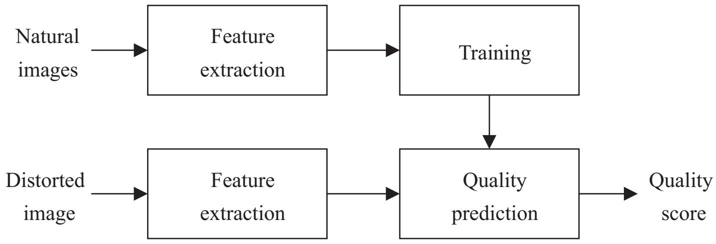


Figure 6 General framework of no-reference image quality assessment algorithms.

subband coefficient probability distributions of the original and distorted images.

Ma et al. [112] used GGD to fit the coefficient distributions of the reorganized DCT subbands. The symmetric city-block distance was employed to measure the image quality. Golestaneh et al. [113] used the adjusted contrast sensitivity function to filter the input image, then the gradient magnitudes were extracted and normalized with entropy of DWT coefficients of the gradient map to calculate the quality. Li et al. [114] used GSM to fit the histogram of image coefficients in the wavelet transform domain. Statistical features extracted from transform domain of reference and distorted images are used to evaluate the quality. Gao et al. [115] used multiscale geometry models to decompose images into different frequency bands. The contrast sensitivity function and just noticeable difference filters are used to mimic the nonlinearities and noticeable variation of human visual system. The histograms of coefficients in different bands are compared to obtain the quality. Zhu et al. [116] utilized a two-level discrete Haar wavelet transform to decompose the input reference and distorted images, and then used sparse representation to extract free-energy-based features from the subband images.

3.3 No-reference IQA algorithms

NR IQA metric aims to predict the quality of image without any information about the original reference image. NR IQA is considered as more challenging than FR and RR IQA because less prior knowledge is used, as illustrated in Figure 6. This also makes NR IQA methods more attractive for practical applications.

On the other hand, although the original image content is not available, assumption of the distortion types can often be made for NR IQA as the application scenario is presumably known for most cases. Therefore, NR IQA methods can be categorized according to distortion types, namely distortion-specific and general-purpose algorithms. Distortion-specific measures generally analyze the artifacts introduced by a specific distortion process and extract relevant features, whereas general-purpose measures can only utilize the general quality features which are supposed to able to describe all types of distortions.

3.3.1 *Distortion-specific NR IQA methods*

If the distortion process is known beforehand, distortion-specific NR IQA measures are favored because of higher accuracy and robustness. The widely studied types of distortions include JPEG compression, blur/noise, and JPEG2000 compression.

(1) JPEG compression. JPEG compression is ubiquitously used in visual communication systems. JPEG compression causes blocking and blurring artifacts at low bitrate, owing to the independent processing of individual blocks. NR IQA measure for JPEG images is one of the most throughly studied area, because of its practical value in quantifying and optimizing existing visual communication systems. Blockiness, as the most prominent JPEG related artifact, is clearly defined and easy to model, and this also facilitates popularity of the study.

The inter-block discontinuity is a representative feature of blocking effect, so many algorithms gauge the JPEG image quality through measuring the difference around block boundaries. Wang et al. [117] modelled the blocky image as a non-blocky image interfered with a pure blocky signal, and the blocking effect is then measured by detecting and estimating the power of the blocky signal. Lee and Park [118] measured the strength of blocking artifacts based on the observation that the pixel values changed abruptly across the boundary, while the pixel values remained unchanged along the entire boundary. Liu and Heynderickx [119, 120] combined the pixel-based distortions, i.e., the blocking artifact, with its local visibility described via a visual masking model. A block grid detector is used to locate the block boundary to reduce the computation complexity. Instead of using pixel discontinuity at the block boundaries, Pan et al. [121] used the edge orientation of the pixels at the block boundaries to measure blockiness. Li et al. [122] evaluated blockiness by measuring the regularities of pseudo structures. The authors also considered the ratio of color-missing blocks because heavier JPEG compression resulted in more color-missing blocks. Li et al. [123] extracted block grids from the image, and then measured the blocking artifacts by quantifying the grid strength and regularity. Min et al. [124] introduced the concept of the most distorted image (MDI), which was derived from the distorted image and suffered from the highest degree of compression. They proposed the pseudo structural similarity (PSS) method for JPEG compressed images by measuring the similarity between pseudo structures of the distorted image and the corresponding MDI.

Besides the inter-block blocking artifacts, intra-block blurring artifacts are also introduced owing to the discard of high frequency DCT coefficients. Some researchers incorporated the intra-block blurring effect into consideration for better precision. Wang et al. [125] predicted JPEG image quality by estimating blockiness described by the average differences across block boundaries, and blurring described by the average absolute difference between in-block image samples and the zero-crossing rate of the difference signal. Perra et al. [126] processed the image with Sobel operator, and measured JPEG image quality by quantifying luminance variation of both the block boundary pixels and the remaining pixels. Zhan and Zhang [127] considered both blockiness along block boundaries and the luminance change within blocks. Gastaldo et al. [128] extracted features derived from the first-order histogram and co-occurrence matrix on a block-by-block basis, and then a circular back-propagation (CBP) neural network [129] was employed for quality regression.

JPEG image quality can also be evaluated from transform domains. Bovik and Liu [130] proposed a measure in the DCT-domain. The blocking artifact is modeled as a 2-D step function by constructing a new block from two adjacent blocks. The block construction and parameter extraction are conducted in the DCT-domain. Chen and Bloom [131] first calculated the absolute difference between adjacent pixels along each column and row. Then a one-dimensional discrete Fourier transform is employed on the difference signal to derive the blockiness measure. Golestaneh and Chandler [132] counted the number of zero-valued DCT coefficients in each block, and the counts are weighted by a quality relevance map, which indicates whether the blocks are natural or generated by JPEG compression. Li et al. [133] employed Tchebichef moments to measure blocking artifacts based on the observation that Tchebichef kernels with different orders are able to capture blockiness. Quantization noise was estimated in [134]. The authors assumed that the probability density function (PDF) of DCT coefficients follows Laplacian distribution, and the PSNR of a given image is predicted by estimating the key parameter of the distribution. Wang et al. [135] extended SSIM to an NR case for DCT-based compressed images through probabilistic models of the quantization noise on spatial and DCT domains.

(2) Blur/noise. Blur and noise are another two common types of distortions widely encountered in practical visual communication systems. Blur can be caused by optical abbreviation, motion, as well as quantization, while noise is almost pervasive in nearly all steps of a visual communication system. We combine the review of NR IQA metrics for blur and noise together because they are actually two opposite types of distortions, i.e., with the loss of useful high frequency contents and the excess of disturbing high frequency contents.

Quality assessment of blur images is a widely studied area and many NR IQA measures have been proposed. Noise is less researched from the perspective of IQA, however, noise estimation for images is an important topic in image processing [136–140] as the noise strength (e.g., σ of a Gaussian model) is a indispensable prior for almost all denoising algorithms. It is noticed that additive noise, as a type of distortion, has little correlation with the image content and therefore its negative impact over perceptual quality is largely proportional with the noise strength [141]. As a consequence, noise estimation result serves as a good quality index for noisy images [142]. Meanwhile, since noise is an opposite type of degradation against blur, many blur measures are found to be effective for noisy images. So in this part we focus on NR blur measures. Because image blurriness and sharpness are two sides of the same coin, the terms of blur metric and sharpness metric are often used interchangeably in the literature.

Early intuitive type of blur measures are based on edge analysis. Marziliano et al. [143] proposed a method by detecting the image edges and analyzing the spread of edges. Ong et al. [144] measured the average extent of image edges. Specifically, the average extent of the slope's spread of an edge in the opposing gradients' directions is measured. Ferzli and Karam [145] integrated the concept of just noticeable blur (JNB) into a probability summation model. JNB indicates the minimum blurring needed to be perceived near the edge given a contrast larger than the just noticeable difference (JND). The authors used a probability summation model to accumulate all blur distortions exceeded the JNB. Narvekar and Karam [146] developed a model based on the human blur perception under varying contrast values. A probabilistic model is employed to estimate the probability of detecting blur at the image edges, then a measure is derived by pooling the cumulative probability of blur detection (CPBD). Feichtenhofer et al. [147] introduced a sharpness measure based on the statistical analysis of local edge gradients.

Some blur metrics also work in spatial domain, yet without explicit edge detection. Bahrami and Kot [148] defined a maximum local variation (MLV) of each pixel as the maximum intensity variation of the pixel with respect to its 8-neighbors. The MLV distribution is analyzed and derived to a sharpness measure. Gu et al. [149] proposed a sharpness measure in autoregressive parameter space. The energy and contrast differences of the estimated local autoregressive coefficients are computed and integrated as the sharpness score. Li et al. [150] proposed a sparse-based sharpness measure. A block-wise sparse representation is first derived, then the sharpness score is computed as the normalized energy computed using the sparse coefficients of a set of high-variance blocks.

Many frequency domain methods are also proposed, since as mentioned, blurring is related to loss of useful high-frequency information. The authors [151] analyzed the histogram of non-zero DCT coefficients, based on the observation that sharp images have high values of alternating current (AC) coefficients. Shaked and Tastl [152] employed localized frequency content analysis in a feature-based context. Vu and Chandler [153] first decomposed the image with a three-level discrete wavelet transform (DWT), and then a weighted average of the log-energies of the DWT subbands was computed as the sharpness. Hassen et al. [154] proposed a local phase coherence (LPC) based sharpness measure in the complex wavelet domain based on the observation that blur introduced degradation of LPC strength near sharp image features. Oh et al. [155] introduced a measure for camera-shaken blur based on spectral statistics, including image spectrum variations across orientations, and some properties of spectral contours of camera shaken images.

There also exist some spatial- and transform-domain hybrid methods. Caviedes and Oberti [156] proposed a measure based on averaged edge profile kurtosis. They detected edges and created an edge profile by assigning them to 8×8 blocks. Then the average 2D kurtosis of the 8×8 DCT blocks is computed. Ciancio et al. [157] fused different sharpness measures and some simple image features into a classifier based on a neural network. A spectral and spatial sharpness (S_3) method is proposed in [158]. S_3 combines a spectral estimator which measures the slope of local magnitude spectrum, with a spatial estimator which measures the total spatial variations. Li et al. [159] extracted multi-scale spatial and spectral features, which are fused through SVR. The spatial features are based on gradient and singular value decomposition, while the spectral features are DCT-domain entropys.

(3) JPEG2000 compression. JPEG2000, as an image format, might not be as successful as expected. However, because JPEG2000 compression is widely included in many IQA databases, and a number of objective quality metrics have been proposed.

Major distortions from JPEG2000 compression include blurring of image details and ringing artifacts around edges. Although many image blur metrics are applicable to JPEG2000 with fairly good results, some researches focused on quantify the ringing artifacts. Marziliano et al. [160] proposed an FR and NR blur metric as well as an FR ringing metric, which are based on the analysis of image edges and adjacent regions. Based on the assumption that natural scenes contain nonlinear dependencies which can be disturbed by the compression process, Sheikh et al. [161] proposed a NSS-based measure by quantifying this disturbance. Sazzad et al. [162, 163] presented a method by fusing pixel distortions estimated from neighboring pixels, and edge information described by zero-crossing rate and histogram features with and without edge preserving filter. Zhang and Le [164] proposed a method based on local pixel activities. Liu et al. [165] first detected the ringing region, then estimated the visibility of ringing artifacts by comparing it with the local background. Liang et al. [166] introduced a ringing metric based on the ringing visibilities of the regions associated with the gradient profiles. Zhang et al. [167] presented a measure based on 1-D and 2-D kurtosis in the DCT domain of the image blocks.

3.3.2 *General-purpose NR IQA methods*

(1) NSS-based methods NSS is a powerful tool for general purpose NR IQA. The motivation is that high quality natural scene images obey some kind of statistical properties, while quality degradations can be deviated from these statistics. A typical NSS-based NR IQA measure consists of 3 key steps: feature extraction, NSS modeling, and feature regression. The features can be extracted from spatial or transform domains. Parametric models such as GGD, multivariate GGD, and asymmetric GGD (AGGD) are used in the NSS modeling step. Finally parameters of the NSS models are regressed to get final quality using support vector machine (SVM) and support vector regression (SVR).

Moorthy and Bovik [168] proposed the blind image quality index (BIQI) through a 2-stage framework involving distortion identification followed by distortion-specific quality assessment. Wavelet transform over three scales and three orientations is conducted, and the subband coefficients are parametrized using a GGD given by

$$f(x|\alpha,\beta,\gamma) = \alpha \mathrm{e}^{-(\beta|x-\mu|)^{\gamma}}, \tag{5}$$

where μ and γ are the mean and shape parameters, α and β are the normalizing and scale parameters. The scene statistics are described through the fitting parameters, which are used as quality features. SVM and SVR are utilized for distortion identification and quality regression. BIQI is further improved to the distortion identification-based image verity and integrity evaluation (DIIVINE) index [169]. The same two-stage framework is utilized, but the quality feature set is enriched to describe the scene statistics more comprehensively by considering the correlations across subbands, scales, and orientations. A learning based blind image quality measure (LBIQ) is developed in [170]. LBIQ extracts features from statistics of complex pyramid wavelet coefficients, texture features described by the cross-scale distribution of coefficient phase, and blur/noise estimation. Principal component analysis (PCA) is performed to reduce the feature dimension and SVM is used to combine these features.

Gao et al. [171] utilized multiple statistical properties in the wavelet domain, including the non-Gaussianity of the wavelet coefficients distribution, the local dependency between the adjacent coefficients, and the exponential decay characteristic of the image energy as the wavelet scale becomes finer, finally multiple kernel learning is used to predict the quality. Zhang et al. [172] presented a complex extension of the DIIVINE (C-DIIVINE), which applies a complex steerable pyramid decomposition to the distorted image, and the corresponding complex-valued subband coefficients are statistically measured as quality features. Wang et al. [173] explored the natural color statistics. Mean subtracted contrast normalized (MSCN) and GGD fitting described above are performed to the color channels of different color spaces. The same quality regression after distortion identification framework as DIIVINE is utilized.

Modeling NSS in the DCT domain enjoys the advantage of lower computational burdens. Saad et al. [174] proposed the BLIINDS index (blind image integrity notator using DCT statistics) using the statistics of local discrete cosine transform coefficients. DCT-based contrast and structure features are extracted as the quality features. Finally the mapping from quality features to quality score is learned through probabilistic prediction models. Saad et al. [175] later introduced the BLIINDS-II index, whose framework is illustrated in Figure 7 [175]. BLIINDS-II applies the GGD described in (5) to model the statistics of the DCT coefficients. The fitted generalized Gaussian model parameters are then used to compute the quality relevant features. Finally, a probabilistic predictive model described by the multivariate GGD

$$f(\boldsymbol{x}|\alpha,\beta,\gamma) = \alpha \mathrm{e}^{-(\beta(\boldsymbol{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\mu))^{\gamma}} \tag{6}$$

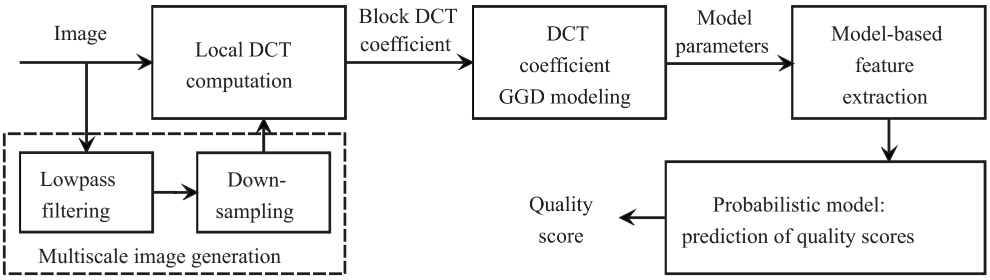


Figure 7 Framework of the NR BLIINDS-II index [175], which is based on the NSS in the DCT domain.

is used to predict the quality score, where Σ is the covariance matrix of the multivariate random variable $\boldsymbol{x}$, and the rest parameters are defined the same as the univariate case.

Spatial domain NSS is also studied for even higher computational efficiency. Mittal et al. [176] proposed the blind/referenceless image spatial quality evaluator (BRISQUE) using the NSS in the spatial domain. BRISQUE models the statistics of the MSCN coefficients

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1}, \tag{7}$$

where i, j are spatial indices, $I(i,j)$ is the given intensity image, $\mu(i,j)$ and $\sigma(i,j)$ are the local mean and variance computed in a local window, respectively. The same GGD described in (5) is used to model the statistics of the MSCN coefficients. Besides that, the products of neighboring MSCN coefficients are modeled through a AGGD:

$$f(x;\gamma,\beta_l,\beta_r) = \begin{cases} \frac{\gamma}{(\beta_l+\beta_r)\Gamma(\frac{1}{\gamma})} \exp(-(\frac{-x}{\beta_l})^{\gamma}), & x < 0, \\ \frac{\gamma}{(\beta_l+\beta_r)\Gamma(\frac{1}{\gamma})} \exp(-(\frac{-x}{\beta_r})^{\gamma}), & x \geqslant 0. \end{cases} \tag{8}$$

The fitting parameters of the GGD and AGGD are extracted as quality features, and regressed through SVR. Based on BRISQUE, Mittal et al. [177] proposed a "completely blind" natural image quality evaluator (NIQE), which uses no human opinion scores for training. The NSS quality features are the same as the BRISQUE, but only image patches with high image contrast are used. Moreover, a multivariate Gaussian model (MVG) which can be described by (6) is used to fit all NSS features. Finally, the distance between the MVGs of high quality images and the target distorted image is computed as the quality.

Zhang et al. [178] developed a feature-enriched version of NIQE named integrated local NIQE (IL-NIQE). Besides the statistics of the MSCN coefficients, IL-NIQE also considers the statistics of gradients, log-Gabor filter responses, and colors. Similarly, MVG is applied to model the statistics of all features, and the distance between the MVGs is computed as the quality. Xue et al. [179] utilized the gradient magnitude (GM) and Laplacian of Gaussian (LOG) features. A joint adaptive normalization is conducted to normalize the GM and LOG features. An index called independency distribution is proposed to measure the joint statistics of them. The marginal distributions and the independency distributions of GM and LOG act as the final quality features, which are integrated through SVR. Lee and Plataniotis [180] proposed the invariance descriptor-based algorithm (IDEAL), which models the statistics of both luminance and color. The luminance statistics are modeled the same way as BRISQUE. The authors also proposed several parametric NSS models to describe the statistical properties of hue, saturation, opponent angle, and spherical angle. The statistics of the luminance and color are used to estimate the quality through SVR. L-moments based statistics are introduced to improve the BRISQUE features in [181]. L-moments make BRISQUE more robust and less sensitive to small variations of NSS. Wu et al. [182] proposed the local pattern statistics index (LPSI), which extracts statistical features described by local binary patterns.

Transform domains other than wavelet are also used. Zhang et al. [183] introduced a measure using the joint statistics of generalized local binary pattern (GLBP). GLBP decomposes the image into multi-scale subband images using the Laplacian of Gaussian (LOG) filters, then joint GLBP histograms extracted

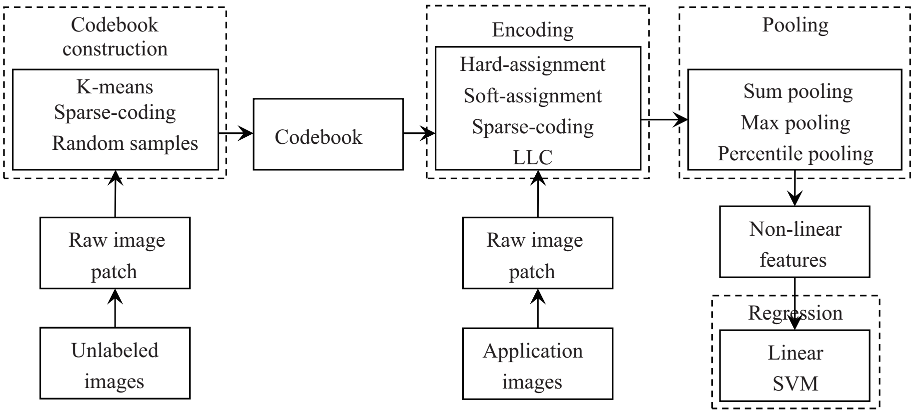


Figure 8 Framework of the NR CORNIA method [189], which is based on unsupervised feature learning.

from the subband images are used as quality features. Lu et al. [184] presented a NSS-based method in contourlet domain. They used a joint distribution to describe the relationship of contourlet coefficients. The statistics of contourlet coefficients are employed as quality features, which are finally combined nonlinearly to the quality score.

Some hybrid methods considering multiple domains are also proposed for better performance. A hybrid no-reference (HNR) model based on a hybrid of curvelet, wavelet, and cosine transforms was proposed in [185]. Natural scene statistics of the log-pdf of the transformed coefficients are modeled to predict the quality. Zhang and Chandler [186] developed a derivative statistics-based quality evaluator (DESIQUE) which extracts NSS features in both the spatial and frequency domains. In spatial domain, log-derivative statistics of single pixels and pixel pairs are modeled, whereas in frequency domain, log-derivative statistics of the log-Gabor filter subband coefficients are modeled. The log-derivative statistics are fitted through GGD, and the fitting parameters are used as the quality features to be regressed to a quality score.

(2) Learning-based methods. With the resurgence of machine learning research, many learning based NR IQA methods were proposed in the last few years. Some methods learn the quality aware features from the images. Ye and Doermann [187,188] introduced visual codebook into the NR IQA problem. They used Gabor filter based local features to construct a visual codebook which is learned from the training images and the correspond quality scores. Then the quality is described by the weighted average of quality scores of codewords. Ye et al. [189] later proposed an unsupervised feature learning framework for NR IQA, namely the CORNIA (codebook representation for no-reference image assessment) method which constructs an unlabeled codebook from whitened raw image patches through K-means clustering. The framework of CORNIA is illustrated in Figure 8. Then soft-assignment coding and feature pooling are performed on a new image to extract the quality features, which are then fused through SVR. Ye et al. [190] later developed a method based on supervised filter learning. They unified the feature extraction and regression processes in a supervised back-projection framework. A quality-aware clustering (QAC) method was introduced in [191]. QAC learns a set of centroids on different quality levels described by FR-IQA measures. The learned centroids are then used as a codebook to infer the quality of image patches. Qualities of all image patches are pooled to the final quality. Several FR measures are combined to generate synthetic quality scores in [192]. The synthetic scores are then used as ground-truth quality scores to replace human opinion scores in training NR-IQA methods to solve the problem of over fitting owing to limited data.

Xu et al. [193] proposed a blind IQA method based on high order statistics aggregation (HOSA). HOSA constructs a codebook in a similar way as CORNIA. Order statistics including mean, variance and skewness are calculated to describe each cluster. Then the local features are assigned to nearest clusters and the differences of between local features and nearest clusters are aggregated to the global quality features. Zhang et al. [194] developed the semantic obviousness metric (SOM) for IQA. SOM detects object-like regions, and then extracts semantic obviousness features on these regions. Local features are extracted from the object-like regions in an unsupervised way which is similar to CORNIA. The semantic and local features are fused through SVM. Mittal et al. [195] applied a topic model on quality-aware visual words described by BRISQUE features, and then examined the topic distributions of the distorted image and pristine natural images to infer the quality. A supervised dictionary learning framework was introduced in [196], where quality scores computed from FR-IQA metrics were used for learning. Specifically, a quality-aware regularization term is added to traditional dictionary learning such that a feature-aware dictionary and a quality-aware dictionary are jointly learned. In [197], local quantized pattern (LQP) is used to extract image features, which are then clustered to construct a codebook. A bag-of-words model is employed for final feature representation, and SVR is used for feature regression. Jiang et al. [198] proposed a codebook based NR IQA method by optimizing multistage discriminative dictionaries, which are learned by performing the label consistent K-SVD algorithm in a stage-by-stage manner.

Sparse coding is also frequently used in NR IQA. He et al. [199] introduced a method based on the sparse representation of natural scene statistics (SRNSS) feature. SRNSS extracts wavelet domain NSS features, which are composed of magnitude, variance, and entropy of the subband wavelet coefficients. Then the features are represented via sparse coding, and weighting ground-truth quality scores by the sparse coding coefficients are computed as the final quality. A dictionary with human opinion scores and hand-crafted features describing two dimensional spatial correlations of images is learned using sparse representation in [200]. The quality is obtained by quantifying the sparse representation coefficients, opinion scores and feature values.

K-nearest neighbor (KNN) is another frequently used approach in NR IQA. Wu et al. [201] proposed a distortion classification and label transfer (TCLT) method, following the two-stage framework used in DIIVINE [169]. The quality features are extracted from multiple domains (DCT, wavelet, and spatial) and multiple color channels (Y, Cb, and Cr). The image's k-nearest-neighbors are searched, and a label transfer method is used to predict the quality. Wu et al. [202] extracted structural features from both spatial-frequency and spatial domains, and a piecewise regression method is employed to train a specific prediction model for each test image using the test image's k-nearest neighbors. A local consistency-aware retriever and an uncertainty-aware evaluator (LOCRUE) is proposed in [203]. LOCRUE searches for similar neighbors of a test image as training set. During searching, the local consistency of the training data is considered to have smoother sample space. The image quality is finally estimated through a sparse Gaussian process. Fang et al. [204] proposed a method to quantify the blurriness, noisiness, and blockiness (BNB) of a image, based on the observation that the adjacent pixels difference follows a generalized Laplace distribution. Three distortion-specific BNB metrics are proposed, and fused using the k-nearest neighbor algorithm.

Rank learning is also a popular tool. Gao et al. [205] utilized learning to rank in NR-IQA. Preference image pairs (PIPs) are generated, and the preference label representing the relative quality of two images is learned using a multiple kernel learning algorithm. A test image is paired with all training images, and the corresponding preference labels are used for quality prediction. A dipIQ method was proposed in [206]. Quality-discriminable image pairs (DIPs) and the corresponding perceptual uncertainty levels are generated. dipIQ employs a neural network based pairwise learning-to-rank algorithm named RankNet [207] to learn the IQA model by incorporating the uncertainty into the loss function. A multi-task rank learning based IQA (MRLIQ) method was developed in [208]. Multiple rank learning based IQA models which are responsible for different distortion types are trained together. The qualities from multiple models are fused using the probabilities of the distortion type to the final quality.

Given the successes of deep neural networks (DNNs) on wide swathes of visual problems [209–214], it is natural to utilize neural network in NR IQA to integrate the extracted quality features or to act as both the feature extractor and regressor. Simple neural network with stacked autoencoders is used in some early studies. A neural network with stacked autoencoders is employed in [215]. The distributions of normalized Y, I and Q channels are used as quality features. The network was pre-trained with greedy layer-wise training, and then fine-tuned through back propagation. A shearlet and stacked autoencoders based no-reference image quality assessment (SESANIA) method was proposed in [216]. SESANIA first performed a shearlet transform, and the sums of subband coefficient amplitudes were calculated as quality features. Stacked autoencoders with a softmax classifier was utilized to mapping features to quality labels. Normalized multi-scale difference of Gaussian (DoG) responses were used as features in [217]. Stacked autoencoders with three hidden layers and a SVM regression were employed for quality regression and pooling.

General regression neural network (GRNN) and deep belief network (DBN) are also useful tool for feature integration. Li et al. [218] utilized a GRNN [219] to fuse the quality features including the mean of phase congruency, the entropy of phase congruency, the entropy and gradient of the distorted image. Tang et al. [220] learned a nonlinear kernel regression function using a rectifier neural network. The utilized neural network was a three-layer DBN [221] pre-trained with unlabeled data and fine-tuned with labeled data. A similar DBN was pre-trained and fine-tuned in [222], but the quality regression step was replaced by a classification framework. The final numerical measurement of image quality was computed as the mean of the quality distribution.

Many NR IQA algorithms benefit from the fast development of CNN. Kang et al. [223] integrated feature learning and regression into a general CNN framework, which consists of one convolutional layer with max and min pooling, two fully connected layers and an output node. The CNN takes normalized image patches as input and outputs the quality score directly. A deeper CNN was used in [93, 224]. A weighted average patch aggregation method was proposed, and the loss was the sum of both the patch-wise loss and the weighted image-wise loss. A compact multi-task CNN was utilized in [225] to estimate image quality and identify distortion type simultaneously in a no-reference setting. Kim and Lee [226] proposed a blind image evaluator based on a convolutional neural network (BIECON). For normalized image patches, CNN model is trained using local patch qualities computed by FR-IQA measures. Then a pooling layer is added to regress the features extracted from CNN, and the whole network is optimized in an end-to-end manner. Gu et al. [227] proposed a vector regression framework for NR IQA by combining belief score estimation and object oriented pooling. First a vector of belief scores is estimated via CNN, then an object oriented pooling is utilized to boost the performance.

Kim et al. [228] utilized a CNN to predict the objective error map, and then predict subjective score. Using the objective error map as an intermediate objective can avoid overfitting problem. Pan et al. [229] designed a deep NR IQA model based on a framework which consisted of a fully CNN and a pooling network. The network was trained using the quality map derived from FR IQA measure. Ma et al. [230] proposed a multi-task end-to-end optimized deep neural network (MEON) for NR IQA. The method consisted of two sub-networks: a distortion identification network and a quality prediction network sharing the early layers. Lin et al. [231] introduced adversarial learning into NR IQA. Liu et al. [232] trained a Siamese network to rank images, and then fine-tuned the network for quality prediction. Talebi et al. [233] trained a CNN on both aesthetic and pixel-level quality datasets, and predicted both mean quality scores and standard deviations. Guan et al. [234] incorporated distortion information, visual importance, and quality perception via a deep neural network. Kim et al. [92] gave an overall introduction on CNN-based IQA.

(3) HVS model based methods. Working mechanism of the HVS, although of which we still have limited knowledge, is generally an important source to design useful features for IQA. Zhai et al. [235] proposed a psychovisual quality metric in free-energy principle. The free-energy principle interprets the perception of an image as an active inference process, in which the brain tries to explain the scene using an internal generative model. The psychovisual quality is closely related to how accurately visual data can be explained by the generative model, which can be quantified using the free energy. A general framework of free energy modeling is illustrated in Figure 9. The readers can refer to [236] for an overview of the free-energy inspired visual quality assessment. A reduced-reference free-energy-based distortion metric (FEDM) and a no-reference free-energy-based quality metric (NFEQM) are developed within the framework in [235]. Gu et al. [237] modified the free-energy-based measure, and proposed an NR free energy-based robust metric (NFERM). NFERM incorporates three groups of features, including the free energy-based feature, some HVS-inspired features such as structural information and gradient magnitude, and spatial NSS features. The three groups of features are regressed using SVR for a final quality

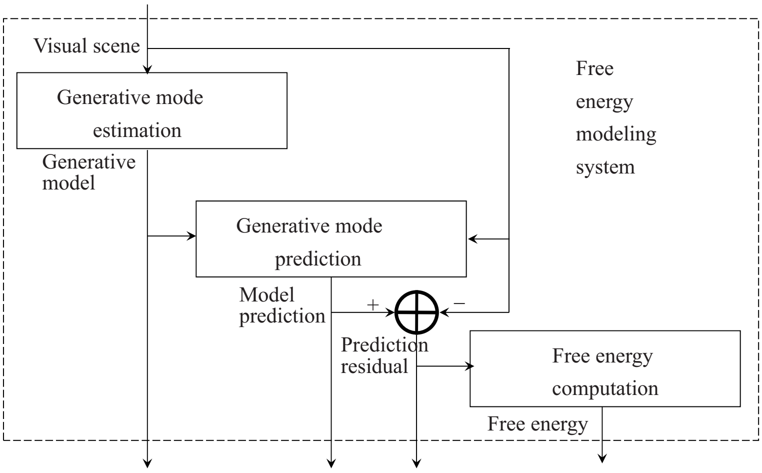


Figure 9 A general framework of free energy modeling [189].

score.

Li et al. [238] proposed an NR-IQA method using structural and luminance features (NRSL) motivated by human visual perception of images. Perceptual structural features described by local binary pattern distribution, and normalized luminance magnitudes distribution are extracted and fused using SVR. Inspired from the HVS's sensitivity to image structural information, Li et al. [239] extracted perceptual features from image first-order and second-order structural patterns, which were described through the distributions of gradient magnitude, and LBPs of the normalized luminance, respectively. Contrary to the conventional IQA metrics, Min et al. [240, 241] utilized a new "reference" called pseudo-reference image (PRI) and introduced a PRI-based blind IQA (BIQA) framework. Different from the traditional reference image, which is assumed to have a perfect quality, PRI is generated from the distorted image and is assumed to suffer from the severest distortion for a given application. Several distortion-specific metrics, one opinion-unaware NR IQA measure, and one opinion-aware NR IQA measure are proposed based on this framework. Wu et al. [242] introduced a pairwise rank-order constraint into the maximum margin regression framework, considering that minimizing average error does not necessarily lead to correct quality rank-orders between the test images. LBP [240–242] was a key quality feature.

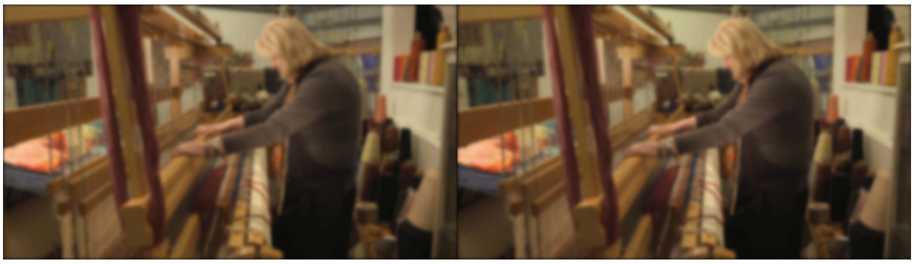
Based on the observation that natural images exhibit redundant information over various scales, Saha et al. [243] used information loss over scales to quantify the distortions. Features extracted from image scale-space, Wavelet domain and Fourier domain are used to formulate the quality score without training. Liu et al. [244] extracted features from wide perceptual domains, such as brightness, contrast, color, distortion, and texture. Then quality score prediction model is built for each feature, and an ensemble method is utilized to combine all quality scores. Freitas et al. [245] utilized the statistics of an orthogonal color plane pattern descriptor to characterize image quality.

4 Objective image quality assessment: emerging topics

This section reviews some emerging topics of IQA in recent years, including stereoscopic IQA, saliency-guided IQA, screen content IQA, Tone-mapping and multiple exposure IQA, retargeting IQA, multi-distortion IQA, authentic distortion IQA, dehazing IQA, virtual reality IQA, and various other emerging topics. We classify those surveyed algorithms in topics/applications, rather than FR, RR and NR types for clearer organization.

4.1 Stereoscopic image quality assessment

Stereoscopic or 3D IQA is popularized with the development of 3D movies and TV programs. The research has both theoretical and practical values as most existing 3D contents and capture and display



(a)

(b)

Figure 10 (Color online) An illustration of 3D distortions, including both left and right views. (a) Symmetrical; (b) asymmetrical. Images are from the Waterloo 3D phase II database [24].

devices still have large room for improvement in terms of visual experience. An illustration of distorted 3D images is given in Figure 10.

4.1.1 *Combination of monocular views and depth sense*

Some researches had shown that the perceived quality of stereoscopic images can be separated into the quality of monocular views and the quality of depth information:

$$\text{O.M.} = \alpha \cdot \text{QI} + \beta \cdot D, \tag{9}$$

where O.M. is the overall measure of stereoscopic image quality, QI is the quality index of monocular views, and D is the term related to perceived depth. Some subjective experiments have been conducted to verify the formulation [246, 247].

Based on the formulation in (9), many 3D IQA metrics have been proposed. Yun et al. [248] combined existing 2D metrics with stereo sense score predicted with the measure of intensity, contrast and structure. Benoit et al. [21] proposed a quality metric for the assessment of stereopairs using the fusion of 2D quality metrics and the depth information. Yang et al. [27] and You et al. [249] followed very similar idea with different choices of quality metrics for both parts in (9). Akhter et al. [250] proposed a metric using segmented monocular image features. Hewage et al. [251] proposed a quality metric for depth map using the extracted edge information. Maalouf et al. [252] considered the color information and contrast sensitivity function (CSF) of human visual system. In [253], a deep neural network is trained to model the process of monocular image quality prediction. Yang et al. [254] extracted 2D wavelet features from monocular images as image content description and 3D features from a depth perception map as depth perception description, which are then fused via deep belief network.

4.1.2 *Binocular visual pathways*

Two visual pathways exist in the binocular visual system [255]. The signals from the left and right eyes will be transformed into uncorrelated or low-correlated signals as the input of visual pathways (channels) [256]. Some stereoscopic image quality assessment algorithms simulate the separate channels in HVS.

(1) Two-eyes pathways. The most direct and intuitive way of analyzing perceptual response is to separate the signal for two eyes. The final quality measure can be obtained by applying 2D image quality assessment algorithms to both the left and right images, and combining these two quality scores into a single number. Moorthy et al. [257] and Gorley et al. [258] noticed that the aggregated two-view quality scores can deliver fairly good performance on symmetrically distorted stereo pairs without computation of the disparity map. Yasakethu et al. [259] averaged the scores of the left and right views measured by a number of 2D quality models with good overall performance. Meegan et al. [260] claimed that the binocular sense of the quality of asymmetric MPEG-2 distorted stereo images is approximately the average of the quality of the two views, and the quality of asymmetric blur distorted stereo images is dominated by the better view. Fang et al. [261] utilized two CNNs to evaluate the qualities of the left and right views, which are then fused via convolutional operations. Zhou et al. [262] introduced a dual-stream interactive network containing left and right view sub-networks for NR stereoscopic IQA.

(2) Frequency bands. In [263], experimental results suggested that the disparity energy signals were integrated across spatial frequency channels for generating a representation of stereoscopic depth in V4. The physiological discoveries suggest that visual signals can be divided into multiple-frequency bands to be evaluated separately:

$$V(I^k) = [V_0^k, \dots, V_{n-1}^k, V_n^k], \tag{10}$$

where V_i^k represents the ith frequency band in the vector $V(I)$ of image I^k, and k can be the right or left view. Lin et al. [264] designed a series of DOG filters of different scales and divided image into multiple-frequency channels. Then the energy of different frequency bands were computed to weight-average toward final results. Jiang et al. [265] considered the binocular interactions by dividing stereoscopic images into frequency components using DOG filters and applying the gain control model to simulate the cyclopean image. The final prediction function is derived from the support vector regression model.

(3) Summation and difference channels. The study in [266], suggested the existence of summation and difference channels for stereopairs:

$$\begin{cases} D = |L - R|, \\ S = \ L + R, \end{cases} \tag{11}$$

where S and D represent the binocular summation and difference of signals separately. Inspired by this, Yang et al. [267] used the summation and difference channels to simulate the visual pathway in human binocular visual system. The combination model using Gabor filter is employed to simulate visual system, and 2D IQA metrics are employed to compare between two combined images. Lin et al. [268] proposed an algorithm considering the influence of distorted disparity information on perceptual quality and the binocular rivalry mechanism of HVS.

(4) Phase and amplitude maps. According to the phase congruency (PC) theory, obvious phase information indicates the orientation where the congruency is maximum over scales [269]. It has been proven in previous researches that simple cells in the primary visual cortex can be well modeled by using Log-Gabor filters [270].

Using the phase congruency theory, Shao et al. [52] considered the JND property and disparity information and further classified images into non-corresponding regions and binocular suppression regions. The log-Gabor filter is employed to calculate phase congruency. Local amplitude and phase are calculated by applying filters of different scales and orientations. Scores are predicted among different regions and are fused into final quality index. Lin et al. [271] estimated the saliency-based visual importance map to modify the amplitude. Besides, binocular visual properties and image low-level features are taken into consideration. There are mainly three parts: the estimation of the saliency-based visual importance map, the computation of cyclopean amplitude and phase maps, and the pooling process.

4.1.3 *Cyclopean images*

There have been some studies towards understanding binocular visual system and the simulation of cyclopean image in the HVS when left-right images are presented to human eyes synchronously [272–274]. Cyclopean image, i.e., the image perceived by binocular stereoscopic vision system with a given image pair, is a popular model in recent research in stereoscopic image quality assessment. Many factors including response of monocular channel, the perceived depth sense, and the interaction mechanism are related to the formation of cyclopean image. One general way to synthesize cyclopean image is

$$\begin{aligned} f_c(I^L, I^R) &= \left(\frac{E^L+1}{E^L+E^R+1}\right)\cdot I^L + \left(\frac{E^R+1}{E^L+E^R+1}\right)\cdot I^R \\ &= g^L\cdot I^L + g^R\cdot I^R, \end{aligned} \tag{12}$$

where E^L and E^R are the energies of stimuli over all the frequency channels for the two views.

With the cyclopean image theory, Liu et al. [275] proposed a strategy to compute the stereoscopic saliency map which was used to weight the cyclopean image. Then the original and distorted cyclopean images are compared by different 2D IQA metrics, and several 2D IQA metrics are compared. Zhou et al. [276] conducted region classification on stereoscopic images to divide two-view images into binocular rivalry and monocular occlusion regions. Then similarities are calculated between original and distorted cyclopean images among these regions, and are integrated into the overall quality score of stereoscopic images by using perceptual modulation function and SVR. Chen et al. [22] developed a framework for assessing the stereoscopic image quality. Monocular quality assessment metrics are employed to predict the synthesized cyclopean image for reference images and distorted images. In another related study [23], part of features were extracted from the formed cyclopean image. Li et al. [277] introduced an adaptive cyclopean image by using ensemble learning.

4.1.4 *Feature extraction*

Features of natural scene statistics are demonstrated to be effective in quality assessment of monocular images. As mentioned in Section 3, there are generally two phases in the quality prediction of 2D images, namely feature extraction and feature integration. Similar strategy can be extended to stereo images with consideration of binocular interactions.

The authors in [278] used GGD to fit the coefficients of luminance subband. Parameters of GGD are estimated as quality features. The final quality is derived by comparing features of the reference and distorted images. Ko et al. [279] proposed an FR stereoscopic IQA system covering a series of parallel scorers. These scorers deal with various distortions and their scores are fused by SVR algorithm to get the final quality measure. Wang et al. [280] proposed an RR 3D IQA method based on the NSS of coefficients of the luminance and disparity maps in the contourlet domain.

Chen et al. [23] used SSIM-based algorithm to estimate the disparity map and uncertainty map and formed the cyclopean image. Based on these maps, features of monocular statistics and stereoscopic statistics are extracted and integrated through SVM. In [281], both spatial-domain and wavelet-domain univariate NSS features, as well as the bivariate and correlation NSS features were extracted and integrated via SVR. Ma et al. [282] found that coefficients of the left and right view images, as well as the disparity map in the reorganized DCT domain had good statistical properties, thus GGD was used to model the distribution of coefficients and the fitted parameters are chosen as quality features.

Zhou et al. [283] simulated the binocular visual perception using binocular energy response and binocular rivalry response. Local patterns of the binocular responses are used to form binocular features, and KNN-learning was used to derive the final quality. In [284], the authors used the log Gabor filter to simulate the responses of left and right V1 cells. Local magnitude, phase and energy patterns were employed to map the image into a histogram for visual content representation. Dictionary and KNN-based learning were used to predict the final quality. In [253] DOG, GM and LBP were used as the basic features to represent textural and structural information of images, deep neural networks are used for feature regression.

4.1.5 *Sparse representation*

As mentioned in Section 3, sparse coding can approximate receptive field of the HVS in response to natural images [285]. It has been widely used in image representation [286] and turned out to be effective for 2D

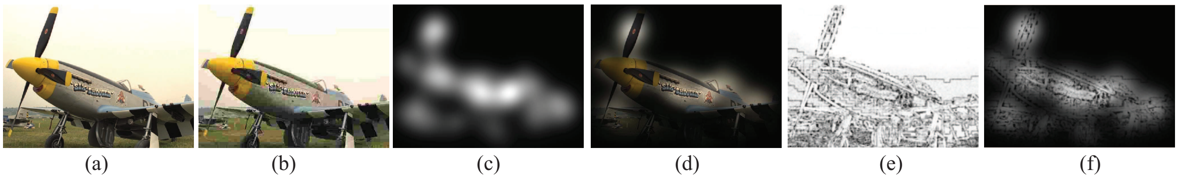


Figure 11 (Color online) An illustration of saliency weighted IQA. (a) Reference image; (b) distorted image; (c) saliency; (d) saliency-weighted image; (e) SSIM; (f) saliency-weighted SSIM.

IQA [80–82, 199, 200]. For stereoscopic images, Shao et al. [287] simulated the disparity considering the spatial frequency, orientations and phase shifts. Then phase-tuned quality lookup and visual codebook are extracted and used to predict the quality. In [288], multi-scale dictionary learning is conducted. Based on the learned dictionary, similarity of the sparse coefficients of the reference and distorted images are calculated and fused to the final quality. Qi et al. [289] represented images via visual primitives, the intra and inter entropys of two views are calculated based on the probability to simulate monocular and binocular cues. Shao et al. [290,291] extracted DOG, HOG and LBP features to construct dictionaries, and SVR was used to train the feature prior model to map features into the final quality. In [292], the authors considered the characteristics of visual fields with the form of dictionary learning and quality lookups. In [293], the authors constructed a database for dictionary learning with labels to indicate whether the difference between two views is noticeable by human eyes. Shao et al. [294] learned modality-specific dictionaries and the corresponding projection matrices from singly distorted images. The reconstruction errors are used to estimate the final quality.

4.2 Saliency-guided IQA

Human visual attention is the behavioral and cognitive process of selectively concentrating on the special region of the stimulus. Image quality and visual attention are two closely related research topics. Image quality is highly related with artifacts located at salient regions, and human attention information can be utilized as weighting map to emphasize salient regions to promote IQA metrics. An illustration of saliency weighted IQA is given in Figure 11. Saliency-guided IQA methods can be classified into 2 categories: subjective visual attention map based weighting and objective saliency map based weighting. It is also noticed that image quality degradation can cause human attention variation especially detected saliency change, thus there are some studies developing IQA metrics by measuring the saliency change.

4.2.1 *Using saliency as weighting map*

(1) Subjective visual attention map based weighting. Eye-tracking data-guided pooling strategy has been explored in recent years to improve performance of IQA metrics. Larson et al. [295] conducted eye-tracking experiments and investigate visual attention change under different image distortion types, severities, and viewing strategies. Liu et al. [296] used eye-tracking data as the weighting information in IQA metrics. They investigated the effectiveness of visual attention data collected under free-viewing and quality-rating tasks, and it was observed that the free-viewing promotes IQA metrics better. Liu et al. [297] also investigated the influence of image content on visual attention data's promotion effect to IQA metrics. They suggested that the images with small inter-observers' attention variations were highly able to profit from saliency pooling. Min et al. [58] collected eye-tracking data for 7 popular IQA datasets, and applied the human fixation data into quality map pooling stage. Wang et al. [298] proposed two novel pooling strategies to incorporate the saliency map into IQA metrics, and they suggested that distortion types affected the performance gain of IQA metrics significantly, and proper pooling strategy should be selected for the specific IQA metric. Rai et al. [299] conducted an eye-tracking experiment in head-mounted display, and proposed a saliency-guided pooling strategy for developing virtual reality IQA metrics. Zhang et al. [300] utilized the eye-tracking data of the distorted images to improve IQA metric. They suggested that the eye-tracking data of both reference image and distorted image were able to promote IQA metrics' performance, and the later showed more encouraging result.

(2) Objective saliency map based weighting. Considering that the eye-tracking devices are generally expensive and are not widely accessible, objective saliency model is also used for large-scale practical IQA tasks. Ma et al. [301] introduced a pooling strategy to apply saliency information into MSSSIM and VIF. They divided the image into overlapped blocks, and calculated the local mean value of each block as weighting coefficient. Zhang et al. [302, 303] used 20 saliency models to weight 12 IQA metrics. Statistical results showed that saliency-guided weighting did promote IQA metrics, and the performance gain was highly related to distortion types and saliency models. Wen et al. [304] proposed an FR IQA metric with saliency weighting, and an objective saliency model using Fourier transform was used for weighting. Xia et al. [305] proposed a saliency weighting method by mimicking human gaze shifting paths via nonnegative matrix factorization. Specifically, they constructed a perceptual space to integrate features, and the simulated gaze shifting paths are refined to train a probabilistic quality evaluation model. Some images contain obvious salient object, while some others lack concentrated salient region. Zhang et al. [306] proposed an efficient saliency dispersion measurement for classifying stimuli. Based on this, they proposed an adaptive method to incorporate saliency into IQA metrics.

Mittal et al. [307] proposed an objective salient region detection algorithm for JPEG distorted images by fusing low level features like contrast, luminance, and quality index. This saliency model is beneficial for developing quality map pooling strategies. Winterlich et al. [308] proposed an NR blur metric for automotive images. They calculated the saliency map using GBVS [309] and Itti&Koch [310] model to weight the quality metric and achieved the better performance. Nasrinpour et al. [311] improved the tone-mapped image quality index (TMQI) [42] by adding saliency-based pooling. The improved method divided the image into small scale patches, and calculated the weighting factor for each patch by AIM [312] saliency model. Similar to [311], Kundu et al. [313] improved the classical TMQI [42] via saliency-based pooling based on Itti&Koch model [310].

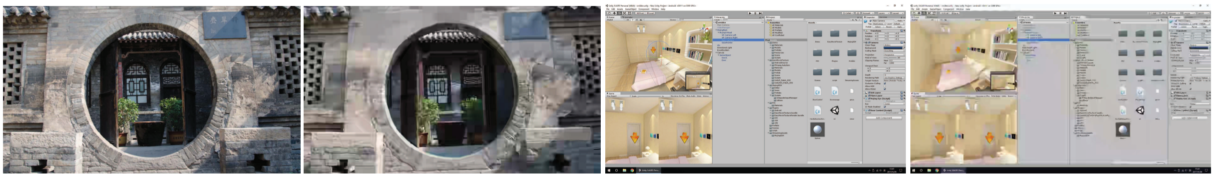
4.2.2 *Using saliency as quality feature*

Instead of being used as a weighting map, visual saliency can be also used as a quality feature. This is possible because various impacting factors can also influence visual attention and image saliency, for example image compression [314], various image transformations and degradations [296, 303, 306, 315, 316], sound [317–320], high-level facial information [321, 322], mental condition and mental healthy [323–325]. In [326], saliency information is used as a feature for NR IQA. It is found that the saliency-guided NSS feature is an efficient descriptor for image quality assessment. Zhang et al. [327] proposed an FR IQA metric named VSI by measuring saliency change. The saliency map which is calculated by the SDSP [328] saliency model is first used as low level quality features, and then serves as a weighting map during the final pooling. Zhang et al. [329] investigated the relationship between distortion-driven gaze pattern variation and image quality degradation. This research suggested that the attention variation caused by distortion is able to predict perceptual quality.

4.3 Screen content IQA

As being computer-generated, SCIs have some characteristics substantially different from NSIs. Therefore, specific IQA measures are designed for SCIs in recent years. A comparison of reference and distorted NSIs and SCIs is given in Figure 12.

Yang et al. [330] constructed a SIQAD database consisting of 20 source and 980 distorted SCIs. An objective measure is also proposed by considering textual and pictorial regions differently. A structure-induced quality metric (SIQM) was proposed in [331], which is described by a SSIM weighted with a structural degradation measurement computed using SSIM as well. Wang et al. [332, 333] presented a method based on visual field adaptation and information content weighting. Ni et al. [37, 334–336] proposed several metrics by measuring the similarity of edge and gradient information. Fu et al. [337] designed a SCI quality model based on multi-scale difference of Gaussian. Gu et al. [338] developed a SCI quality measure using saliency-guided GM similarity. Similarly, the structural variation based quality index (SVQI) proposed in [339] models the global and local structure variations introduced by distortions.



(a) (b) (c) (d)

Figure 12 (Color online) A comparison of reference and distorted NSIs and SCIs. (a) Reference NSI; (b) distorted NSI; (c) reference SCI; (d) distorted SCI. Images are from the CCT database [38].

Fang et al. [340] introduced an uncertainty weighting method to fuse the quality of textual and pictorial regions. Zhang et al. [341] introduced an FR SCI quality measure based on separate measures of the edge-structure degradations for different segmented regions which are segmented and classified via CNN.

Wang et al. [342] constructed a compressed SCI quality assessment database, and proposed an RR measure for SCIs via learning a set of wavelet domain features. An RR SCI measure was proposed in [343] by combining statistical features extracted from the primary visual information and the amount of uncertainty. Jakhetiya et al. [344] proposed an RR IQA algorithm for SCIs based on a perceptually relevant prediction model which emphasizes more on the textual regions. Gu et al. [345] learned a blind quality evaluator for SCIs based on an FR SCI quality evaluator. A CNN-based framework was used in [346]. Gu et al. [347] developed a blind metric via learning specific features and fusing through SVR. And the regression model was trained with an FR SCI quality measure. Shao et al. [348] presented a blind measure for SCIs from a perspective of sparse representation. Fang et al. [349] proposed an NR SCI quality measure by incorporating the statistical luminance and texture features.

Besides those quality measures specifically designed for SCIs, some measures are proposed for multiple types of images. Ye et al. [190] presented a general-purpose NR IQA measure working for both NSIs and document images captured from document files. Xu et al. [193] proposed a blind IQA method based on HOSA. HOSA works for multiple image types including NSI, SCI, and document image. Min et al. [124] proposed an NR quality measure for JPEG compressed NSIs and SCIs. Min et al. [38] constructed a CCT database which contained reference and compressed NSIs, CGIs, and SCIs. They also proposed a unified content-type adaptive (UCA) blind IQA model that was applicable across content types. Zhou et al. [350] designed a blind quality measure for SCIs and NSIs based on a dictionary of learned local and global quality features.

4.4 Tone-mapping IQA

High dynamic range (HDR) images need to be converted low dynamic range (LDR) images to be visualized on standard LDR displays. Tone-mapping operators (TMOs) are used in this step. An illustration of tone-mapped images is given in Figure 13. To assess the perceptual quality of the tone-mapped images created by different TMOs, many quality metrics are proposed.

Yeganeh and Wang [42] introduced a TMQI by combining a multi-scale signal fidelity measure modified from SSIM and a naturalness measure based on natural scene intensity statistics. Ma et al. [351] improved TMQI to TMQI-II through modifying the two key components in TMQI, i.e., the structural fidelity and statistical naturalness measures. Gu et al. [352, 353] developed a blind tone-mapped quality index (BTMQI) by analyzing the information, naturalness and structure of the tone-mapped images. Nafchi et al. [354] proposed a feature similarity index for tone-mapped images (FSITM). FSITM compared the locally weighted mean phase angle maps of the HDR and LDR images. Visual attention was considered in [311], and a saliency weighted tone-mapped quality index (STMQI) was proposed. Kundu et al. [355] conducted a large-scale crowdsourced study for tone-mapped HDR images. The constructed database includes images obtained by both TMOs and multi-exposure fusion (MEF) algorithms, with and without postprocessing. Based upon the subjective study, Kundu et al. [355] later designed NR HDR picture quality models using space-domain NSS features and HDR-specific gradient-based features. Hadizadeh et al. [356] extracted perceptually relevant quality-related features such as structural fidelity, naturalness, and overall brightness, and integrated the features to the tone-mapping quality. Yue et al. [357] introduced

Figure 13 (Color online) Tone-mapped images derived from different TMOs. Images are from the TMID [42].

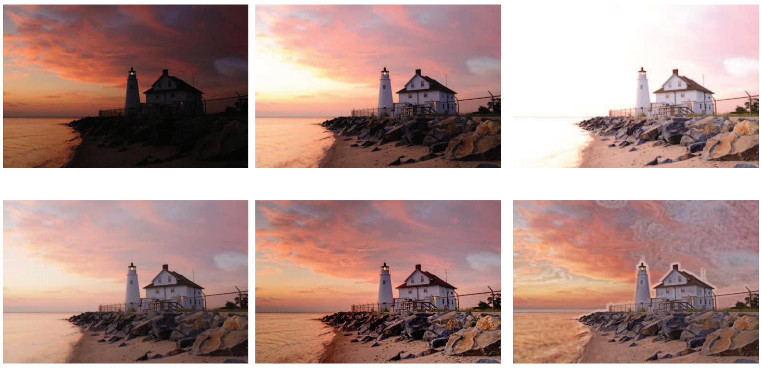


Figure 14 (Color online) Images with multiple exposure levels (top) and the fused images (bottom). Images are from the MEF database [44].

a biologically inspired tone-mapping quality measure from the perspective of color information processing in the brain. Yue et al. [358] integrated colorfulness, naturalness and structure features to assess the tone-mapping quality.

4.5 Multi-exposure fusion IQA

MEF combines images taken with multiple exposure levels for an informative output. An illustration of images with multiple exposure levels and the fused images is given in Figure 14. Some IQA models are proposed for the MEF evaluation.

Xydeas and Petrovic [359] proposed a measure of the visual information transferred from the input images into the fused image. Qu et al. [360] designed an information measure for image fusion evaluation based on mutual information. Piella and Heijmans [361] utilized local measures to estimate how well the salient information was preserved in the fused images. Cvejic et al. [362] proposed an image fusion metric based on mutual information and Tsallis entropy. Chen and Varshney [363] developed a human perception inspired quality metric for image fusion based on regional information. Zheng et al. [364] proposed a ratio of spatial frequency error (rSFe) measure based on the spatial frequency which reflects the local intensity variation. Wang and Liu [365] proposed an edge information preservation inspired metric. Hossny et al. [366] modified the classical method based on mutual information [360]. Chen and Blum [367] computed the quality of a fused image from a contrast preservation map which described the relationship between the fused image and each source image. Hassen et al. [368] developed an objective quality measure for multi-exposure multi-focus image fusion. The developed fusion quality index (FQI) incorporates several key factors including contrast preservation, sharpness, and structural preservation. Ma et al. [44] assessed the quality of MEF images based on the multi-scale SSIM principle and a measure of patch structural consistency. Kundu et al. [355] conducted subjective and objective quality study of images obtained by TMOs and MEF algorithms.

4.6 Retargeting IQA

Image regargeting usually crops image and changes resolution and aspect ratio, without other obvious distortions. An illustration of original and retargeted images is given in Figure 15. Some IQA methods

Figure 15 (Color online) Original (the 1st column) and retargeted images (the 2nd–4th columns). Images are from the MIT RetargetMe [31].

are designed specifically for this type of operation.

In [369], SIFT flow and optical flow are used to calculate the correspondence between images. The loss of area in salient regions, the preservation of aspect ratio of salient objects and the degradations of local shapes are extracted as quality features and fused through SVR. In [370], features of image pairs were fed into the pairwise rank learning to learn a ranking model. Image quality scores are generated by referring to an exponential curve fitting function based on the rankings. Zhang et al. [371] solved the backward registration problem to reveal the geometric change during image retargeting. Information loss and visual distortion are considered to assess local block changes. The overall retargeting quality is predicted via visual importance based pooling. Fang et al. [372] created an image retargeting quality database, and predicted the perceptual quality of retargeted images by weighting the SSIM map using the saliency map. Ma et al. [32] also built an image retargeting quality database. Several quality metrics were evaluated on the database, and experimental results demonstrated that the performance will be better if the shape distortion and content information loss were considered. In [373], SIFT flow was used to estimate the dense correspondence. The estimated SIFT-flow vector and the calculated saliency map were employed to measure the geometric distortion. The information loss was also calculated with the saliency map.

Chen et al. [374] designed a bidirectional natural salient scene distortion model for image retargeting quality assessment. The model consisted of a image NSS measurement, a salient global structure distortion measurement, and a bidirectional salient information loss measurement. Zhang et al. [375] utilized multiple-level features for retargeting quality assessment, including low-level aspect ratio similarity, mid-level edge group similarity, and high-level face block similarity. Zhang et al. [376] predicted the retargeting quality by evaluating the retargeting fidelity and detecting deformation inconsistency on three levels: region-level segmentation, patch-level partition and pixel-level correspondence. Liang et al. [377] designed a retargeting quality measure based on five factors: preservation of salient regions, analysis of the influence of artifacts, preservation of the global structure of the image, compliance with well-established aesthetics rules, and preservation of symmetry.

4.7 Multiple distortions

In a visual communication system, the steps of capturing, processing, compression, transmission, and decompression, all introduce different distortions in a consecutive manner. Therefore, it is an important topic to study IQA for multiple distortions. An illustration of images degraded by multiple distortions is given in Figure 16.

Gu et al. [34] created a multiply distorted image database and proposed a blind measure inspired by free energy based brain theory. The metric incorporated the measures of single distortions and the joint effects between different distortion types. Zhang et al. [378] used cartoon-texture decomposition to separate image into cartoon part with salient edges and texture part with noises. Then the degrees of

Figure 16 (Color online) Images degraded by multiple distortions. Images are from the MDID2013 [34].

Figure 17 (Color online) A comparison of simulated and authentic distortions. Images are from the LIVE [13] and LIVE Wild [40] databases.

blur and noise are estimated from different parts. The joint effect is also measured. A pooling strategy is employed to derive the final quality score. Lu et al. [379] extracted NSS features used in SRNSS, BRISQUE-L and BIQI which are representative and fast for calculation. The SRNSS and BIQI features are both extracted in the wavelet domain while the BRISQUE-L features are extracted in the spatial domain. An improved bag-of-words (BOW) model is applied to encode extracted features, and the SVR is used to map image features to the quality score. In [380], the authors described images by feature maps calculated from color Gaussian jet of the image. LBPs are applied to measure the structural degradations on these feature maps. In [381], LBPs were calculated on the gradient map of images to measure the structural degradation caused by multiple distortions. The histogram of LBP is further weighted by gradient magnitude. SVR is adopted to map features to the quality index. Zhang et al. [382] proposed an opinion-unaware blind IQA method for multiply and singly distorted images. They first identified the distortion types, and then estimated the distortion parameters for quality prediction.

4.8 Authentic distortions

Almost all of the distortions used in IQA research are simulated with explicit models. However, in more realistic settings, the image acquisition, processing and transmission steps can introduce so called authentic distortions that cannot be clearly modeled. A comparison of simulated and authentic distortions is given in Figure 17. Some IQA algorithms for authentic distortions are proposed in recent years.

Brooks et al. [383] considered realistic distortions from compression and error concealment in video compression/transmission applications and simulated typical distortions encountered in other applications. A complex wavelet SSIM (CWSSIM) model was proposed considering viewing distance. Yang et al. [384] proposed a blind image quality assessment model. Features including the statistical property and characteristics for authentic distortions were extracted by considering the perception of human visual system. Support vector regressor was trained to predict the final score. Ghadiyaram et al. [385, 386] extracted NSS features in different color spaces and transform domains, then SVR or deep belief net (DBN) was used to fuse all features. Li et al. [238] extracted structural features by analyzing LBP descriptors and properties of luminance magnitudes in the form of histogram, and features at different scales were fused via SVR to generate a final score. Liu et al. [387] extracted the quality-aware features from the low-level human vision characteristics and the high-level brain activities in free-energy principle to predict the camera image quality. Besides authentically distorted images, authentically distorted videos were also studied in [388].

4.9 View synthesis IQA

View synthesis can be of great value in various applications, such as virtual reality, augmented reality, free viewpoint video, and light field. Various view synthesis methods have been proposed in recent years, and

Figure 18 (Color online) DIBR synthesized images. Images are from the IRCCyN/IVC DIBR image database.

Figure 19 (Color online) Hazy and dehazed images. Images are from the DHQ database.

these methods can introduce different visual artifacts and result in imperfect quality. Quality evaluation of DIBR techniques has been widely researched. DIBR tries to synthesize a view for any viewpoint given the sparse views and the scene depth. An illustration of DIBR synthesized images was given in Figure 18.

Bosc et al. [45] constructed the IRCCyN/IVC DIBR images database, which is the basis of the following DIBR-synthesized image quality evaluation studies. Battisti et al. [389] proposed a 3D synthesized view image quality metric (3DSwIM) by comparing the statistical features of wavelet subbands of two input images. Sandić-Stanković et al. [390, 391] proposed a morphological pyramid peak signal-to-noise ratio (PSNR) metric (MP-PSNR), a morphological wavelet PSNR (MW-PSNR) metric, and the reduced versions of MP-PSNR and MW-PSNR. Li et al. [392] proposed a metric by measuring local geometric artifacts and global sharpness. Gu et al. [393] introduced a metric based on the local image descriptor auto-regressor. Tian et al. [394] evaluated the quality by measuring the typical synthesis distortions. Zhou et al. [395] assessed the DIBR-synthesized video quality by measuring temporal flickering. Ling et al. [396] proposed a quality measure for free-viewpoint videos by quantifying the elastic changes of multi-scale motion trajectories.

4.10 Dehazing IQA

Dehazing IQA is driven by the need of evaluating the effect of various image dehazing outputs. Example hazy and dehazed images were given in Figure 19. In the literature, image dehazing can be evaluated using two strategies: using synthetic hazy images, or using real hazy images. When using synthetic hazy images, the ground-truth haze-free images are available, and it is an FR IQA problem. While if real hazy images are used, there are no ground-truth haze-free images, thus it is an NR IQA problem.

In the strategy of using synthetic hazy images, generally FR IQA measures like PSNR and SSIM are used for quality evaluation [397]. However, in [47], the authors constructed a synthetic haze removing quality (SHRQ) database, and found that traditional FR IQA measures like PSNR and SSIM are not effective enough for synthetic haze removing quality evaluation. Thus they introduced an FR dehazing quality measure based on some existing FR IQA methodologies. Besides evaluation using synthetic hazy

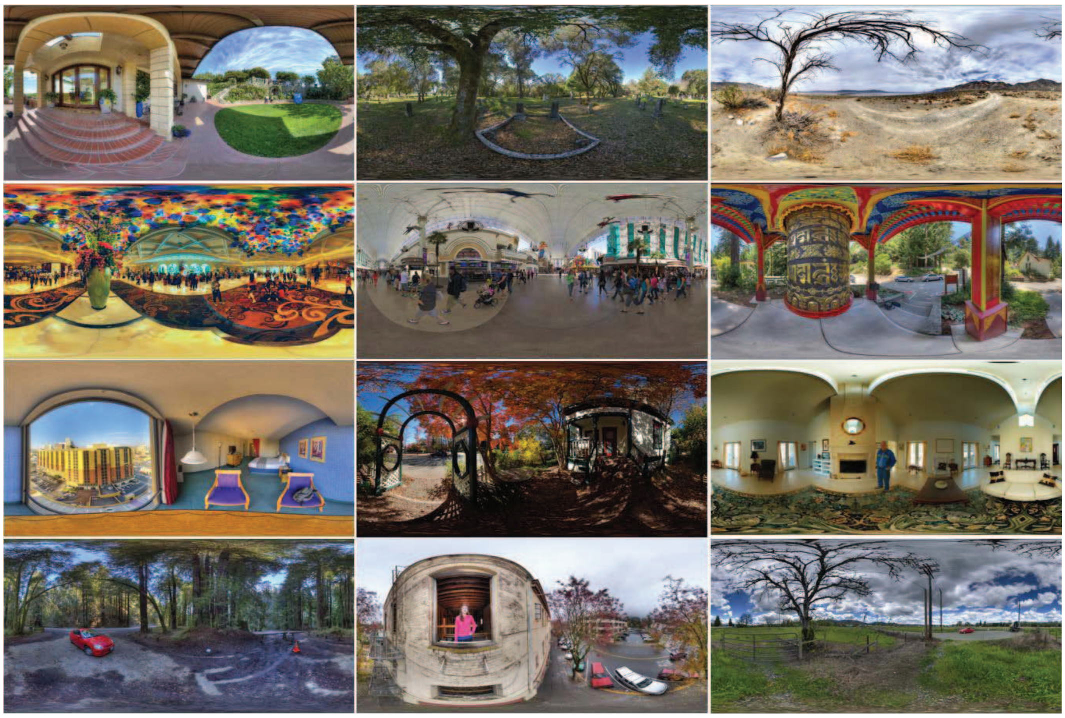


Figure 20 (Color online) Example VR images from the OIQA database.

images, evaluation using real hazy images is also desirable. In [48], the authors conducted a subjective dehazing quality assessment study and constructed the IVC dehazed image database, in which most of the hazy images were realistic hazy images and a few were synthetic hazy images. However, no objective quality measure is proposed. Min et al. [46] constructed the dehazing quality (DHQ) database (all of the hazy images are real hazy images) and proposed an objective dehazing quality index (DHQI) by extracting and fusing three groups of features, including: haze-removing features, structure-preserving features, and over-enhancement features. These 3 groups of features have captured the most key aspects of dehazing. Besides the above studies, some early studies also utilized some simple image contrast or edge descriptors to evaluate the dehazing effect [398], and sometimes the general-purpose IQA measures like BRISQUE [176] and NIQE [177] will also be used.

4.11 Virtual reality IQA

Virtual reality (VR) or omnidirectional IQA has also attracted much attention recently, especially in the recent 5 years. Example VR images are shown in Figure 20. It is driven by the prevalence of various VR applications. Owing to the specific characteristics of omnidirectional viewing experience in the head mounted display (HMD) and various specific VR content processing techniques, quality assessment of VR image needs to be specifically studied.

Several VR IQA databases have been constructed to facilitate the relevant research, including the OIQA Database [49], CVIQ Database [50, 53], and the LIVE 3D VR IQA Database [51]. Besides these open VR IQA databases which are publicly available, there are also some subjective VR IQA studies in the literature, for example assessment of encoded omnidirectional videos [399, 400], and assessment of visually induced motion sickness in immersive videos [401]. While in the context of objective VR IQA, a multichannel CNN for blind 360-degree image quality assessment (MC360IQA) was proposed in [50, 53]. In this method, each 360-degree image is decomposed into six viewport images, which are evaluated by six channels of CNNs. There are also some methods which generalize traditional IQA methods like PSNR and SSIM for VR contents, for example the sphere based PSNR (S-PSNR) [402], the weighted spherical PSNR (WS-PSNR) [403], the Craster parabolic projection PSNR (CPP-PSNR) [404], the non-content-based perceptual PSNR (NCPPSNR) and content-based perceptual PSNR (CP-PSNR) [400]. Another

VR content perception topic which is closely related VR quality assessment is visual attention prediction in VR. Visual attention prediction can be of great value in various VR content processing techniques, for example VR quality assessment, compression, and transmission [405]. A grand challenge for this topic was organized at ICME'17 and some VR saliency models were proposed [406–412]. Specifically, head movement prediction, eye movement prediction, and scanpath prediction were investigated in these studies. An overview of the VR perception, assessment and compression studies was given in [413].

4.12 Other relevant topics

Besides the topics which have been reviewed above, there are also many other quality assessment topics dedicated to the assessment of various types and formats of multimedia content. Limited by space, we will not give an overview for each of these topics, but we will give a summary of these research topics and several representative studies for each topic below.

- Video quality assessment. Video quality assessment studies are not included in this study. A survey of video quality assessment was given in [414]. The most common category of video quality measures are based on content analysis, and they are effective for the quality evaluation of some common distortions, for example, video compression, blur, and noise [415, 416]. Besides that, there are also some methods which predict video quality based on video specification parameters [417, 418] and some other system parameters [419].
- Video streaming quality assessment. This category of studies address distortions that occur over longer time spans, for example bitrate changes and rebuffering events [420–425].
- Quality assessment of audio-visual signals. Considering that most of the quality assessment studies focused on only single-modality video or audio signals, some studies were conducted to evaluate the joint quality of audio-visual signals [426, 427].
- Quality assessment of compressed authentically distorted image. This topic aims at predicting the quality of compressed authentically distorted images which are captured in the wild [428].
- Quality assessment of contrast changed images. This topic focuses specifically on contrast changed images, because most of the traditional IQA measures are not effective for this kind of distortion [429–433].
- Image aesthetic assessment. Different from traditional IQA which aims to evaluate the quality degradation introduced by various image distortions like image compression, blur, and noise, image aesthetic assessment tries to assess photo quality based on photographic rules, and its influencing factors may include uses of lighting, contrast, color, and image composition [434–436]. The image aesthetic measures are also very different from image quality measures. An overview of this topic was given in [434].
- Transparently encrypted image quality assessment. Transparent encryption tries to protect partial content and fulfill the security and quality requirements, and some quality measures are proposed to evaluate the quality of such process [437, 438].
- Light field quality assessment. Owing to the complex structures and high dimensions of light field data, conventional processing techniques for 2D images may not work well for light field, thus light field quality assessment is specifically researched [439–441].
- Quality assessment of X-Ray images or TeraHertz (THz) images. Besides visible light IQA, the quality assessment of X-Ray images [442] or TeraHertz (THz) images [443, 444] have also aroused some attention recently.
- Quality assessment of sonar images. The degradations of sonar images are mainly introduced during the process of acquisition and transmission, and some quality measures are proposed for the corresponding quality evaluation [445, 446].
- Quality assessment considering the environment. Besides the quality of the content, the conditions of the environment in which the human beings view the content also have large impacts on the quality of experience. Some studies have investigated the influence of such environment conditions, for example, backlight luminance was considered in [447] while the environment luminance was further considered in [448].

Table 2 Performance of full-reference image quality assessment algorithms

Types	Metrics	CSIQ		LIVE		TID2008		TID2013	
		SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
Spatial method	PSNR	0.8058	0.7512	0.8756	0.8723	0.5531	0.5223	–	–
	SSIM [59]	0.8756	0.8612	0.9479	0.9449	0.7749	0.7732	–	–
	VPSNR [450]	0.96	0.98	–	–	0.92	0.96	–	–
	IW-SSIM [61]	0.9213	0.9144	0.9567	0.9522	0.8559	0.8579	–	–
	MS-SSIM [60]	0.9133	0.8991	0.9513	0.9489	0.8542	0.8451	–	–
	NSER [70]	0.9337	0.9473	0.9419	0.9395	0.7404	0.7957	–	–
	GS [66]	0.8005	0.7998	0.8756	0.8723	0.5794	0.5726	–	–
	VGS [68]	0.9662	0.9692	0.9696	0.9686	0.8983	0.9079	–	–
	IGM [63]	0.9401	0.9280	0.9580	0.9578	0.8902	0.8858	–	–
	GMSD [67]	0.957	0.954	0.960	0.960	0.891	0.879	–	–
	SVCM [69]	0.951	0.949	0.964	0.962	0.874	0.889	0.787	0.857
	IFS [451]	0.9581	0.9576	0.9599	0.9586	0.8903	0.8810	0.8697	0.8791
Transformation-based	VSNR [19]	0.8109	0.7355	0.9271	0.9229	0.7046	0.6820	–	–
	VIF [95]	0.9195	0.9277	0.9632	0.9598	0.7496	0.8090	–	–
	MIQE [96]	0.911	0.916	0.964	0.962	0.807	0.840	–	–
	MIQEC [96]	0.930	0.926	0.961	0.960	0.788	0.829	–	–
	DCT-QM [99]	0.9332	0.7674	0.9557	0.8260	0.8392	0.6641	0.8544	0.6791
	SC-DM [100]	0.9423	0.7863	0.9475	0.8092	0.9021	0.7252	0.9003	0.7270
	SC-QI [100]	0.9434	0.7870	0.9480	0.8098	0.9051	0.7294	0.9052	0.7327
	IQDM [98]	0.9058	0.8976	0.9336	0.9536	0.8415	0.8369	–	–
Learning-based	Q [75]	0.881	0.900	0.925	0.924	0.817	0.816	–	–
	SFF [80]	0.9627	0.9643	0.9649	0.9632	0.8767	0.8817	–	–
	ParaBoost [77]	0.9733	0.9766	0.9819	0.9802	0.9772	0.9767	0.9575	0.9567
	QASD [81]	0.9516	0.9466	0.9646	0.9602	0.8899	0.8877	0.8657	0.8894
	KRR [82]	0.9141	0.9197	0.9574	0.9587	0.8865	0.8903	0.7969	0.8220
	NMF [452]	0.9727	0.9763	0.9760	0.9758	0.9466	0.9513	–	–
	DeepSim [90]	0.919	0.919	0.974	0.968	–	–	0.846	0.872
	DADF [79]	–	–	–	–	0.930	0.782	–	–
Other	FSIM [64]	0.9242	0.9120	0.9634	0.9597	0.8805	0.8738	–	–
	Li et al. [97]	0.933	0.928	0.946	0.936	0.861	0.869	–	–

5 Evaluation of IQA models

5.1 Evaluation protocol

According to the recommendations given by video quality experts group (VQEG) [449], which are widely accepted in the research of IQA, performance of objective assessment algorithms can be evaluated from three aspects: prediction accuracy, prediction monotonicity and prediction consistency. For performance comparison, VQEG suggests to reduce the nonlinearity of the prediction values of objective models. A five parameter monotonic logistic function is commonly used to map the computed quality scores:

$$p = \beta_1 \left(0.5 - \frac{1}{1 + \mathrm{e}^{\beta_2(o-\beta_3)}}\right) + \beta_4 o + \beta_5, \tag{13}$$

where o and p are the computed and mapped scores. After the nonlinear mapping, the following five evaluation criteria can be used to measure the performance of the IQA models.

- Spearman rank order correlation coefficient (SROCC):

$$\mathrm{SROCC} = 1 - \frac{6\sum_{n=1}^{N} d_i^2}{N(N^2-1)}, \tag{14}$$

where d_i represents the difference between the i-th images's ranks in subjective and objective evaluations, and N is the number of testing images.

Table 3 Performance of reduced-reference image quality assessment algorithms

Metrics	CSIQ		LIVE		TID2008		TID2013	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
RRED [110]	–	–	0.9429	0.9385	–	–	–	–
RR-SSIM [111]	0.8527	0.8426	0.9129	0.9194	0.7210	0.7231	–	–
Wu et al. [102]	–	–	0.732	0.725	0.528	–	–	–
WNISM [112]	–	–	0.880	0.883	–	–	–	–
REDLOG [113]	0.8576	0.8560	0.9455	0.9372	0.6864	0.7326	0.6829	0.7400
DNT marginal [114]	–	–	0.9287	0.9173	–	–	–	–
SPCRM-SCHARR [453]	0.8889	0.8906	0.9131	0.9097	0.7567	0.7403	–	–

Table 4 Performance of no-reference image quality assessment algorithms

Types	Metrics	CSIQ		LIVE		TID2008		TID2013	
		SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
NSS-based	BIQI [168]	–	–	0.8195	0.8205	–	–	–	–
	DIIVINE [169]	–	–	0.916	0.917	–	–	–	–
	BLIINDS-II [175]	–	–	0.9202	0.9232	–	–	–	–
	BRISQUE [176]	–	–	0.9395	0.9424	–	–	–	–
	NIQE [177]	–	–	0.9135	0.9147	–	–	–	–
	IL-NIQE [178]	0.822	0.865	0.902	0.906	–	–	0.521	0.648
	GM-LOG [179]	0.9243	0.9457	0.9511	0.9551	0.9369	0.9406	–	–
	IDEAL [180]	0.8683	0.8913	0.9409	0.9462	–	–	0.7190	0.7674
	DESIQUE [186]	0.928	0.942	0.9437	0.9465	0.919	0.925	–	–
	C-DIIVINE [172]	0.910	0.935	0.9444	0.9474	0.921	0.925	–	–
	GLBP [183]	–	–	0.921	–	–	–	–	–
	LPSI [182]	–	–	0.9511	0.9542	0.9399	–	–	–
Learning-based	CORNIA [189]	–	–	0.942	0.935	0.813	0.837	–	–
	HOSA [193]	0.9298	0.9480	0.9504	0.9527	–	–	0.9521	0.9592
	QAC [191]	0.8627	0.8768	0.8857	0.8608	0.8697	0.8377	–	–
	SRNSS [199]	–	–	0.9304	0.9318	–	–	–	–
	SOM [194]	–	–	0.964	0.962	–	–	–	–
	TCLT [201]	0.891	–	0.934	0.935	0.872	–	–	–
	LQP [197]	0.9109	0.9255	0.9289	0.9307	–	–	0.9244	0.9325
	BNB [204]	–	–	0.9508	0.9497	–	–	–	–
	NRHC [78]	–	–	0.8776	0.8714	–	–	–	–
	PIPs [205]	0.843	–	0.938	–	–	–	0.779	–
	DIPs [206]	0.930	0.949	0.958	0.957	–	–	0.877	0.894
	MRLIQ [208]	0.9219	–	0.9528	–	–	–	–	–
	SESANIA [216]	–	–	0.9340	0.9476	0.8936	0.9069	–	–
	BIECON [226]	–	–	0.961	0.962	0.923	–	–	–
HVS-based	NFERM [237]	0.9142	–	0.9405	0.9457	0.9156	–	–	–
	NRSL [238]	0.930	0.954	0.952	0.956	–	–	0.945	0.959
	BSD [239]	0.9330	0.9489	0.9618	0.9653	0.9557	0.9673	–	–
	DIQES [243]	0.8561	0.8879	0.8966	0.9034	0.8223	0.8103	0.8355	0.8348

Table 5 Performance of 3D image quality assessment algorithms

Types		Metrics	NBU 3D		LIVE 3D Phase-I		LIVE 3D Phase-II		MCL-3D	
			SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
Binocular visual pathways	FR	Liu et al. [275]	0.9206	0.9330	0.9336	0.9459	0.9030	0.9162	–	–
	FR	SDM-GSSIM [267]	–	–	0.9248	0.9332	–	–	–	–
	FR	Lin et al. [268]	–	–	0.9256	0.9391	0.9196	0.9292	–	–
	FR	Lin et al. [271]	–	–	0.9314	0.9366	0.8824	0.8984	–	–
Cyclopean images	FR	NN-MS-SSIM [275]	–	–	0.9385	0.9371	–	–	–	–
	NR	Zhou et al. [276]	–	–	0.887	0.928	0.823	0.861	–	–
Feature extraction	NR	Zhou et al. [454]	–	–	0.904	0.934	0.890	0.905	–	–
	NR	Zhou et al. [284]	–	–	0.901	0.929	0.819	0.856	0.837	0.867
	FR	STRIQE [278]	–	–	0.9223	0.9275	0.8920	0.9019	–	–
	NR	S3D-BLINQ [281]	–	–	–	–	0.905	0.913	–	–
	RR	Wang et al. [280]	–	–	0.8890	0.8921	–	–	–	–
	RR	Ma-1 et al. [282]	–	–	0.9034	0.9033	0.8093	0.8431	–	–
	RR	Ma-2 et al. [282]	–	–	0.9052	0.9056	0.7938	0.8179	–	–
	FR	3D-DQE [253]	0.9420	0.9493	0.9449	0.9565	0.9106	0.9265	0.9040	0.9138
Sparse representation	NR	Shao et al. [287]	0.9026	0.9061	0.8756	0.9042	–	–	–	–
	FR	Shao et al. [288]	0.9411	0.9413	0.9251	0.9350	0.8494	0.8628	–	–
	RR	Qi et al. [289]	–	–	–	–	0.867	0.915	–	–
	NR	Shao et al. [290]	0.9375	0.9486	0.9440	0.9531	0.8849	0.9034	–	–
	NR	Shao et al. [291]	0.9305	0.9479	0.9498	0.9572	–	–	–	–
	NR	Shao et al. [293]	–	–	0.8667	0.8846	0.8717	0.9095	–	–
	NR	NUMBLIM [294]	0.8757	0.8679	0.8849	0.8913	0.8054	0.7843	–	–

- Kendall rank order correlation coefficient (KROCC):

$$\mathrm{KROCC} = \frac{N_c - N_d}{0.5(N-1)N}, \tag{15}$$

where N_c and N_d express the numbers of concordant and discordant pairs in the testing data, respectively.

- Pearson linear correlation coefficient (PLCC):

$$\mathrm{PLCC} = \frac{\sum_{i=1}^{N}(p_i - \bar{p})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{N}(p_i - \bar{p})^2(s_i - \bar{s})^2}}, \tag{16}$$

where s_i and p_i indicate the i-th image's subjective score and converted objective rating after nonlinear mapping, $\bar{s}$ and $\bar{p}$ are the mean of all s_i and p_i.

- Root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - s_i)^2}. \tag{17}$$

- Mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|p_i - s_i|. \tag{18}$$

Among the above five evaluation criteria, SROCC and KROCC measure the prediction monotonicity, PLCC estimates the linearity and consistency, RMSE and MAE evaluate the prediction accuracy. Higher SROCC, KROCC, PLCC scores and lower RMSE, MAE scores indicate better correlation with the subjective ratings.

5.2 Performance comparison

We compare performance of those surveyed IQA methods in this subsection. Not all reviewed metrics are included because not all algorithms are publicly available. Considering the randomness in both the algorithms and the evaluation processes, for a fair comparison, we only record the performance reported in the original papers.

Performances of FR, RR, NR, and 3D IQA measures are listed in Table 2 [19, 59–61, 63, 66–70, 75, 77, 79–82, 90, 95–100, 450–452], Table 3 [102, 110–114, 453], Table 4 [78, 168, 169, 172, 175–180, 182, 183, 186, 189, 191, 193, 194, 197, 199, 201, 204–206, 208, 216, 226, 237–239, 243], and Table 5 [52, 253, 264, 267, 271, 275, 276, 278–282, 284, 287–291, 293, 294, 454], respectively. SRCC and PLCC performance are used owing to limit of space and because of their popularity.

General-purpose FR, RR and NR IQA measures are compared using LIVE [13], CSIQ [16], TID2008 [14], and TID2013 [15], which are four widely used general-purpose IQA databases. 3D IQA measures are compared on LIVE 3D Phase I [22], LIVE 3D Phase II [23], Ningbo 3D [52], and MCL-3D [28] databases. Note that most FR and NR IQA measures are training-free, and are tested on the whole IQA databases. And many NR IQA measures involve training, mostly with an 80% train – 20% test strategy. Those performance results can be used as a reference for designing quality metric and optimization algorithm for visual communication systems.

6 Conclusion

In this survey, we conducted a comprehensive and up-to-date review of perceptual image quality assessment. Subjective quality assessment databases including classical and emerging ones, were reviewed first. Then traditional full-reference, reduced-reference and no-reference image quality metrics were analyzed in sequel. Emerging topics in the field such as stereoscopic images, saliency guided approach, screen content images, tone mapping and multiple expose image, were also reviewed. Finally, performance evaluation model were introduced and the performance of many quality metrics were listed and compared. This survey serves as an overview of the quality assessment problem for visual communication for researcher in related areas. It also facilitates researchers within the field finding solutions and trends in their study.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61901260, 61831015, 61521062, 61527804).

References

1 Wang Z, Bovik A C. Mean squared error: love it or leave it? a new look at signal fidelity measures. IEEE Signal Process Mag, 2009, 26: 98–117
2 Wang Z, Bovik A C. Reduced- and no-reference image quality assessment. IEEE Signal Process Mag, 2011, 28: 29–40
3 Lin W, Kuo C C J. Perceptual visual quality metrics: a survey. J Visual Commun Image Represent, 2011, 22: 297–312
4 Moorthy A K, Bovik A C. Visual quality assessment algorithms: what does the future hold? Mult Tools Appl, 2011, 51: 675–696
5 Chandler D M. Seven challenges in image quality assessment: past, present, and future research. ISRN Signal Process, 2013, 2013
6 He L, Gao F, Hou W, et al. Objective image quality assessment: a survey. Int J Comput Math, 2014, 91: 2374–2388
7 Mohammadi P, Ebrahimi-Moghadam A, Shirani S. Subjective and objective quality assessment of image: a survey. 2014. ArXiv: 14067799
8 Manap R A, Shao L. Non-distortion-specific no-reference image quality assessment: a survey. Inf Sci, 2015, 301: 141–160
9 Xu S, Jiang S, Min W. No-reference/blind image quality assessment: a survey. IETE Tech Rev, 2017, 34: 223–245
10 Winkler S. Analysis of public image and video databases for quality assessment. IEEE J Sel Top Signal Process, 2012, 6: 616–625
11 Winkler S, Subramanian R. Overview of eye tracking datasets. In: Proceedings of IEEE International Workshop on Quality of Multimedia Experience, 2013. 212–217
12 Series B. Methodology for the Subjective Assessment of the Quality of Television Pictures. Recommendation ITU-R BT, 2012. 500–13

13 Sheikh H R, Wang Z, Cormack L, et al. LIVE image quality assessment database release 2. http://live.ece.utexas.edu/research/quality
14 Ponomarenko N, Lukin V, Zelensky A, et al. TID2008-a database for evaluation of full-reference visual quality assessment metrics. Adv Modern Radioelectron, 2009, 10: 30–45
15 Ponomarenko N, Jin L, Ieremeiev O, et al. Image database TID2013: peculiarities, results and perspectives. Signal Process Image Commun, 2015, 30: 57–77
16 Larson E, Chandler D. Consumer subjective image quality database. 2009. http://visionokstateedu/csiq/
17 Le Callet P, Autrusseau F. Subjective quality assessment IRCCyN/IVC database. http://ivc.univ-nantes.fr/en/databases/Subjective_Database/
18 Horita Y, Shibata K, Kawayoke Y, et al. Mict image quality evaluation database. http://mict.eng.u-toyama.ac.jp/mictdb.html
19 Chandler D M, Hemami S S. VSNR: a wavelet-based visual signal-to-noise ratio for natural images. IEEE Trans Image Process, 2007, 16: 2284–2298
20 Ma K, Duanmu Z, Wu Q, et al. Waterloo exploration database: new challenges for image quality assessment models. IEEE Trans Image Process, 2017, 26: 1004–1016
21 Benoit A, Le Callet P, Campisi P, et al. Quality assessment of stereoscopic images. EURASIP J Image Video Process, 2009, 2008: 659024
22 Chen M J, Su C C, Kwon D K, et al. Full-reference quality assessment of stereopairs accounting for rivalry. Signal Process Image Commun, 2013, 28: 1143–1155
23 Chen M J, Cormack L K, Bovik A C. No-reference quality assessment of natural stereopairs. IEEE Trans Image Process, 2013, 22: 3379–3391
24 Wang J, Rehman A, Zeng K, et al. Quality prediction of asymmetrically distorted stereoscopic 3D images. IEEE Trans Image Process, 2015, 24: 3400–3414
25 Wang X, Yu M, Yang Y, et al. Research on subjective stereoscopic image quality assessment. In: Proceedings of the SPIE, 2009
26 Zhou J, Jiang G, Mao X, et al. Subjective quality analyses of stereoscopic images in 3DTV system. In: Proceedings of IEEE International Conference on Visual Communications and Image Processing, 2011. 1–4
27 Yang J, Hou C, Zhou Y, et al. Objective quality assessment method of stereo images. In: Proceedings of 3DTV Conference: the True Vision-Capture, Transmission and Display of 3D Video, 2009. 1–4
28 Song R, Ko H, Kuo C C J. MCL-3D: a database for stereoscopic image quality assessment using 2d-image-plus-depth source. J Inf Sci Eng, 2015, 31: 1593–1611
29 Jung Y J, Sohn H, Lee S I, et al. Predicting visual discomfort of stereoscopic images using human attention model. IEEE Trans Circ Syst Video Tech, 2013, 23: 2077–2082
30 Goldmann L, de Simone F, Ebrahimi T. Impact of acquisition distortion on the quality of stereoscopic images. In: Proceedings of International Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2010
31 Rubinstein M, Gutierrez D, Sorkine O, et al. Retarget me a benchmark for image retargeting. ACM Trans Graph, 2010, 29: 1–9
32 Ma L, Lin W, Deng C, et al. Image retargeting quality assessment: a study of subjective scores and objective metrics. IEEE J Sel Top Signal Process, 2012, 6: 626–639
33 Jayaraman D, Mittal A, Moorthy A K, et al. Objective quality assessment of multiply distorted images. In: Proceedings of the 46th Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2012. 1693–1697
34 Gu K, Zhai G, Yang X, et al. Hybrid no-reference quality metric for singly and multiply distorted images. IEEE Trans Broadcast, 2014, 60: 555–567
35 Sun W, Zhou F, Liao Q. MDID: a multiply distorted image database for image quality assessment. Pattern Recogn, 2017, 61: 153–168
36 Yang H, Fang Y, Lin W, et al. Subjective quality assessment of screen content images. In: Proceedings of IEEE International Workshop on Quality of Multimedia Experience, 2014. 257–262
37 Ni Z, Ma L, Zeng H, et al. ESIM: edge similarity for screen content image quality assessment. IEEE Trans Image Process, 2017, 26: 4818–4831
38 Min X, Ma K, Gu K, et al. Unified blind quality assessment of compressed natural, graphic, and screen content images. IEEE Trans Image Process, 2017, 26: 5462–5474
39 Gu K, Xu X, Qiao J, et al. Learning a unified blind image quality metric via on-line and off-line big training instances. IEEE Trans Big Data, 2019. doi: 10.1109/TBDATA.2019.2895605
40 Ghadiyaram D, Bovik A C. Massive online crowdsourced study of subjective and objective picture quality. IEEE Trans Image Process, 2016, 25: 372–387
41 Virtanen T, Nuutinen M, Vaahteranoksa M, et al. CID2013: a database for evaluating no-reference image quality assessment algorithms. IEEE Trans Image Process, 2015, 24: 390–402
42 Yeganeh H, Wang Z. Objective quality assessment of tone-mapped images. IEEE Trans Image Process, 2013, 22: 657–667
43 Kundu D, Ghadiyaram D, Bovik A C, et al. Large-scale crowdsourced study for tone-mapped HDR pictures. IEEE Trans Image Process, 2017, 26: 4725–4740
44 Ma K, Zeng K, Wang Z. Perceptual quality assessment for multi-exposure image fusion. IEEE Trans Image Process, 2015, 24: 3345–3356
45 Bosc E, Pepion R, Le Callet P, et al. Towards a new quality metric for 3-D synthesized view assessment. IEEE J Sel

Top Signal Process, 2011, 5: 1332–1343
46 Min X, Zhai G, Gu K, et al. Objective quality evaluation of dehazed images. IEEE Trans Intell Trans Syst, 2019, 20: 2879–2892
47 Min X, Zhai G, Gu K, et al. Quality evaluation of image dehazing methods using synthetic hazy images. IEEE Trans Mult, 2019, 21: 2319–2333
48 Ma K, Liu W, Wang Z. Perceptual evaluation of single image dehazing algorithms. In: Proceedings of IEEE International Conference on Image Processing, 2015. 3600–3604
49 Duan H, Zhai G, Min X, et al. Perceptual quality assessment of omnidirectional images. In: Proceedings of IEEE International Symposium on Circuits and Systems, 2018. 1–5
50 Sun W, Min X, Zhai G, et al. MC360IQA: a multi-channel CNN for blind 360-degree image quality assessment. IEEE J Sel Top Signal Process, 2020, 14: 64–77
51 Chen M, Jin Y, Goodall T, et al. Study of 3D virtual reality picture quality. 2019. ArXiv: 191003074
52 Shao F, Lin W, Gu S, et al. Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. IEEE Trans Image Process, 2013, 22: 1940–1953
53 Sun W, Luo W, Min X, et al. Mc360iqa: the multi-channel cnn for blind 360-degree image quality assessment. In: Proceedings of IEEE International Symposium on Circuits and Systems, 2019. 1–5
54 Liu H, Heynderickx I. Studying the added value of visual attention in objective image quality metrics based on eye movement data. In: Proceedings of IEEE International Conference on Image Processing, 2009. 3097–3100
55 Alers H, Redi J, Liu H, et al. Studying the effect of optimizing image quality in salient regions at the expense of background content. J Electron Imag, 2013, 22: 043012
56 Redi J A, Liu H, Zunino R, et al. Interactions of visual attention and quality perception. In: Proceedings of SPIE, 2011. 7865: 78650S
57 Engelke U, Maeder A, Zepernick H J. Visual attention modelling for subjective image quality databases. In: Proceedings of IEEE International Workshop on Multimedia Signal Processing, 2009. 1–6
58 Min X, Zhai G, Gao Z, et al. Visual attention data for image quality assessment databases. In: Proceedings of IEEE International Symposium on Circuits and Systems, 2014. 894–897
59 Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process, 2004, 13: 600–612
60 Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment. In: Proceedings of IEEE Asilomar Conference on Signals, Systems, and Computers, 2003. 1398–1402
61 Wang Z, Li Q. Information content weighting for perceptual image quality assessment. IEEE Trans Image Process, 2011, 20: 1185–1198
62 Tan H L, Li Z G, Tan Y H, et al. A perceptually relevant MSE-based image quality metric. IEEE Trans Image Process, 2013, 22: 4447–4459
63 Wu J J, Lin W S, Shi G M, et al. Perceptual quality metric with internal generative mechanism. IEEE Trans Image Process, 2013, 22: 43–54
64 Zhang L, Zhang L, Mou X Q, et al. FSIM: a feature similarity index for image quality assessment. IEEE Trans Image Process, 2011, 20: 2378–2386
65 Gu K, Li L, Lu H, et al. A fast reliable image quality predictor by fusing micro- and macro-structures. IEEE Trans Ind Electron, 2017, 64: 3903–3912
66 Liu A M, Lin W S, Narwaria M. Image quality assessment based on gradient similarity. IEEE Trans Image Process, 2012, 21: 1500–1512
67 Xue W, Zhang L, Mou X, et al. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. IEEE Trans Image Process, 2014, 23: 684–695
68 Zhu J Y, Wang N C. Image quality assessment by visual gradient similarity. IEEE Trans Image Process, 2012, 21: 919–933
69 Zhan Y, Zhang R, Wu Q. A structural variation classification model for image quality assessment. IEEE Trans Mult, 2017, 19: 1837–1847
70 Zhang M, Mou X, Zhang L. Non-shift edge based ratio (NSER): an image quality assessment metric based on early vision features. IEEE Signal Process Lett, 2011, 18: 315–318
71 Capodiferro L, Jacovitti G, Di Claudio E D. Two-dimensional approach to full-reference image quality assessment based on positional structural information. IEEE Trans Image Process, 2012, 21: 505–516
72 Di Claudio E D, Jacovitti G. A detail-based method for linear full reference image quality prediction. IEEE Trans Image Process, 2018, 27: 179–193
73 Ding L, Huang H, Zang Y. Image quality assessment using directional anisotropy structure measurement. IEEE Trans Image Process, 2017, 26: 1799–1809
74 Sun W, Liao Q, Xue J H, et al. SPSIM: a superpixel-based similarity index for full-reference image quality assessment. IEEE Trans Image Process, 2018, 27: 4232–4244
75 Narwaria M, Lin W S. Objective image quality assessment based on support vector regression. IEEE Trans Neural Netw, 2010, 21: 515–519
76 Shnayderman A, Gusev A, Eskicioglu A M. An SVD-based grayscale image quality measure for local and global assessment. IEEE Trans Image Process, 2006, 15: 422–429
77 Liu T J, Liu K H, Lin J Y, et al. A paraboost method to image quality assessment. IEEE Trans Neural Netw Learn Syst, 2017, 28: 107–121

78 He L, Wang D, Liu Q, et al. Fast image quality assessment via supervised iterative quantization method. Neurocomputing, 2016, 212: 121–127

79 Peng P, Li Z N. General-purpose image quality assessment based on distortion-aware decision fusion. Neurocomputing, 2014, 134: 117–121

80 Chang H W, Yang H, Gan Y, et al. Sparse feature fidelity for perceptual image quality assessment. IEEE Trans Image Process, 2013, 22: 4007–4018

81 Li L, Cai H, Zhang Y, et al. Sparse representation-based image quality index with adaptive sub-dictionaries. IEEE Trans Image Process, 2016, 25: 3775–3786

82 Yuan Y, Guo Q, Lu X. Image quality assessment: a sparse learning way. Neurocomputing, 2015, 159: 227–241

83 Ahar A, Barri A, Schelkens P. From sparse coding significance to perceptual quality: a new approach for image quality assessment. IEEE Trans Image Process, 2018, 27: 879–893

84 Pang Y, Sun M, Jiang X, et al. Convolution in convolution for network in network. IEEE Trans Neural Netw Learn Syst, 2017, 29: 1587–1597

85 Pang Y, Cao J, Wang J, et al. JCS-Net: joint classification and super-resolution network for small-scale pedestrian detection in surveillance images. IEEE Trans Inform Forensic Secur, 2019, 14: 3322–3331

86 Zhao C R, Chen K, Zang D, et al. Uncertainty-optimized deep learning model for small-scale person re-identification. Sci China Inf Sci, 2019, 62: 220102

87 Chen J, Lian Z H, Wang Y Z, et al. Irregular scene text detection via attention guided border labeling. Sci China Inf Sci, 2019, 62: 220103

88 Liu B, Chen X, Han Y, et al. Accelerating DNN-based 3D point cloud processing for mobile computing. Sci China Inf Sci, 2019, 62: 212206

89 Zhu J, Zeng H, Jin X, et al. Joint horizontal and vertical deep learning feature for vehicle re-identification. Sci China Inf Sci, 2019, 62: 199101

90 Gao F, Wang Y, Li P, et al. DeepSim: deep similarity for image quality assessment. Neurocomputing, 2017, 257: 104–114

91 Wang H, Fu J, Lin W, et al. Image quality assessment based on local linear information and distortion-specific compensation. IEEE Trans Image Process, 2017, 26: 915–926

92 Kim J, Lee S. Deep learning of human visual sensitivity in image quality assessment framework. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017. 1969–1977

93 Bosse S, Maniry D, Muller K R, et al. Deep neural networks for no-reference and full-reference image quality assessment. IEEE Trans Image Process, 2018, 27: 206–219

94 Sheikh H R, Bovik A C, Veciana G D. An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Trans Image Process, 2005, 14: 2117–2128

95 Sheikh H R, Bovik A C. Image information and visual quality. IEEE Trans Image Process, 2006, 15: 430–444

96 Demirtas A M, Reibman A R, Jafarkhani H. Full-reference quality estimation for images with different spatial resolutions. IEEE Trans Image Process, 2014, 23: 2069–2080

97 Li S, Zhang F, Ma L, et al. Image quality assessment by separately evaluating detail losses and additive impairments. IEEE Trans Mult, 2011, 13: 935–949

98 Tang C, Yang X, Zhai G. Image quality/distortion metric based on α-stable model similarity in wavelet domain. J Visual Commun Image Represent, 2014, 25: 1746–1757

99 Bae S H, Kim M. DCT-QM: a DCT-based quality degradation metric for image quality optimization problems. IEEE Trans Image Process, 2016, 25: 4916–4930

100 Bae S H, Kim M. A novel image quality assessment with globally and locally consilient visual quality perception. IEEE Trans Image Process, 2016, 25: 2392–2406

101 Redi J A, Gastaldo P, Heynderickx I, et al. Color distribution information for the reduced-reference assessment of perceived image quality. IEEE Trans Circ Syst Video Technol, 2010, 20: 1757–1769

102 Wu J, Lin W, Shi G, et al. Reduced-reference image quality assessment with visual information fidelity. IEEE Trans Mult, 2013, 15: 1700–1705

103 Decherchi S, Gastaldo P, Zunino R, et al. Circular-ELM for the reduced-reference assessment of perceived image quality. Neurocomputing, 2013, 102: 78–89

104 Bampis C G, Gupta P, Soundararajan R, et al. SpEED-QA: spatial efficient entropic differencing for image and video quality. IEEE Signal Process Lett, 2017, 24: 1333–1337

105 Zhang Y, Phan T D, Chandler D M. Reduced-reference image quality assessment based on distortion families of local perceived sharpness. Signal Process Image Commun, 2017, 55: 130–145

106 Min X, Gu K, Zhai G, et al. Saliency-induced reduced-reference quality index for natural scene and screen content images. Signal Process, 2018, 145: 127–136

107 Liu Y, Zhai G, Gu K, et al. Reduced-reference image quality assessment in free-energy principle and sparse representation. IEEE Trans Mult, 2018, 20: 379–391

108 Gao X, Lu W, Li X, et al. Wavelet-based contourlet in quality evaluation of digital images. Neurocomputing, 2008, 72: 378–385

109 Wang Z, Wu G X, Sheikh H R, et al. Quality-aware images. IEEE Trans Image Process, 2006, 15: 1680–1689

110 Soundararajan R, Bovik A C. RRED indices: reduced reference entropic differencing for image quality assessment. IEEE Trans Image Process, 2012, 21: 517–526

111 Rehman A, Zhou Wang A. Reduced-reference image quality assessment by structural similarity estimation. IEEE

Trans Image Process, 2012, 21: 3378–3389
112 Ma L, Li S, Zhang F, et al. Reduced-reference image quality assessment using reorganized DCT-based image representation. IEEE Trans Mult, 2011, 13: 824–829
113 Golestaneh S A, Karam L J. Reduced-reference quality assessment based on the entropy of DWT coefficients of locally weighted gradient magnitudes. IEEE Trans Image Process, 2016, 25: 5293–5303
114 Li Q, Wang Z. Reduced-reference image quality assessment using divisive normalization-based image representation. IEEE J Sel Top Signal Process, 2009, 3: 202–211
115 Gao X B, Lu W, Tao D C, et al. Image quality assessment based on multiscale geometric analysis. IEEE Trans Image Process, 2009, 18: 1409–1423
116 Zhu W, Zhai G, Min X, et al. Multi-channel decomposition in tandem with free-energy principle for reduced-reference image quality assessment. IEEE Trans Mult, 2019, 21: 2334–2346
117 Wang Z, Bovik A C, Evan B. Blind measurement of blocking artifacts in images. In: Proceedings of IEEE International Conference on Image Processing, 2000. 3: 981–984
118 Lee S, Park S J. A new image quality assessment method to detect and measure strength of blocking artifacts. Signal Process Image Commun, 2012, 27: 31–38
119 Liu H, Heynderickx I. A no-reference perceptual blockiness metric. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. 865–868
120 Liu H, Heynderickx I. A perceptually relevant no-reference blockiness metric based on local image characteristics. EURASIP J Adv Signal Process, 2009, 2009: 263540
121 Pan F, Lin X, Rahardja S, et al. Using edge direction information for measuring blocking artifacts of images. Multidim Syst Sign Process, 2007, 18: 297–308
122 Li L, Lin W, Zhu H. Learning structural regularity for evaluating blocking artifacts in JPEG images. IEEE Signal Process Lett, 2014, 21: 918–922
123 Li L, Zhou Y, Wu J, et al. GridSAR: grid strength and regularity for robust evaluation of blocking artifacts in JPEG images. J Visual Commun Image Represent, 2015, 30: 153–163
124 Min X, Zhai G, Gu K, et al. Blind quality assessment of compressed images via pseudo structural similarity. In: Proceedings of IEEE International Conference on Multimedia and Expo, 2016. 1–6
125 Wang Z, Sheikh H R, Bovik A C. No-reference perceptual quality assessment of JPEG compressed images. In: Proceedings of IEEE International Conference on Image Processing, 2002. 477–480
126 Perra C, Massidda F, Giusto D D. Image blockiness evaluation based on sobel operator. In: Proceedings of IEEE International Conference on Image Processing, 2005. 389
127 Zhan Y, Zhang R. No-reference JPEG image quality assessment based on blockiness and luminance change. IEEE Signal Process Lett, 2017, 24: 760–764
128 Gastaldo P, Parodi G, Redi J, et al. No-reference quality assessment of JPEG images by using CBP neural networks. In: Proceedings of International Conference on Artificial Neural Networks, 2007. 564–572
129 Ridella S, Rovetta S, Zunino R. Circular backpropagation networks for classification. IEEE Trans Neural Netw, 1997, 8: 84–97
130 Bovik A C, Liu S. DCT-domain blind measurement of blocking artifacts in DCT-coded images. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. 3: 1725–1728
131 Chen C, Bloom J A. A blind reference-free blockiness measure. In: Proceedings of Pacific-Rim Conference on Multimedia, 2010. 112–123
132 Golestaneh S A, Chandler D M. No-reference quality assessment of JPEG images via a quality relevance map. IEEE Signal Process Lett, 2014, 21: 155–158
133 Li L, Zhu H, Yang G, et al. Referenceless measure of blocking artifacts by tchebichef kernel analysis. IEEE Signal Process Lett, 2014, 21: 122–125
134 Ci W, Dong H, Wu Z, et al. Example-based objective quality estimation for compressed images. IEEE MultiMedia, 2019. doi: 10.1109/MMUL.2009.77
135 Wang C, Shen M, Yao C. No-reference quality assessment for DCT-based compressed image. J Visual Commun Image Represent, 2015, 28: 53–59
136 Liu C, Freeman W T, Szeliski R, et al. Noise estimation from a single image. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006. 901–908
137 Zoran D, Weiss Y. Scale invariance and noise in natural images. In: Proceedings of IEEE International Conference on Computer Vision, 2009. 2209–2216
138 Zhai G, Wu X. Noise estimation using statistics of natural images. In: Proceedings of IEEE International Conference on Image Processing, 2011. 1857–1860
139 Tang C, Yang X, Zhai G. Noise estimation of natural images via statistical analysis and noise injection. IEEE Trans Circ Syst Video Technol, 2015, 25: 1283–1294
140 Dong L, Zhou J, Tang Y Y. Effective and fast estimation for image sensor noise via constrained weighted least squares. IEEE Trans Image Process, 2018, 27: 2715–2730
141 Zhai G, Wu X. On monotonicity of image quality metrics. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2012. 1157–1160
142 Zhai G, Kaup A, Wang J, et al. A dual-model approach to blind quality assessment of noisy images. APSIPA Trans Signal Inf Process, 2015, 4: 29–32
143 Marziliano P, Dufaux F, Winkler S, et al. A no-reference perceptual blur metric. In: Proceedings of IEEE Interna-

tional Conference on Image Processing, 2002
144 Ong E, Lin W, Lu Z, et al. A no-reference quality metric for measuring image blur. In: Proceedings of International Symposium on Signal Processing and its Applications, 2003. 469–472
145 Ferzli R, Karam L J. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). IEEE Trans Image Process, 2009, 18: 717–728
146 Narvekar N D, Karam L J. A no-reference image blur metric based on the cumulative probability of blur detection (CPBD). IEEE Trans Image Process, 2011, 20: 2678–2683
147 Feichtenhofer C, Fassold H, Schallauer P. A perceptual image sharpness metric based on local edge gradient analysis. IEEE Signal Process Lett, 2013, 20: 379–382
148 Bahrami K, Kot A C. A fast approach for no-reference image sharpness assessment based on maximum local variation. IEEE Signal Process Lett, 2014, 21: 751–755
149 Gu K, Zhai G T, Lin W S, et al. No-reference image sharpness assessment in autoregressive parameter space. IEEE Trans Image Process, 2015, 24: 3218–3231
150 Li L, Wu D, Wu J, et al. Image sharpness assessment by sparse representation. IEEE Trans Mult, 2016, 18: 1085–1097
151 Marichal X, Ma W Y, Zhang H. Blur determination in the compressed domain using DCT information. In: Proceedings of IEEE International Conference on Image Processing, 1999. 386–390
152 Shaked D, Tastl I. Sharpness measure: towards automatic image enhancement. In: Proceedings of IEEE International Conference on Image Processing, 2005. 937
153 Vu P V, Chandler D M. A fast wavelet-based algorithm for global and local image sharpness estimation. IEEE Signal Process Lett, 2012, 19: 423–426
154 Hassen R, Wang Zhou, Salama M M A. Image sharpness assessment based on local phase coherence. IEEE Trans Image Process, 2013, 22: 2798–2810
155 Oh T, Park J, Seshadrinathan K, et al. No-reference sharpness assessment of camera-shaken images by analysis of spectral structure. IEEE Trans Image Process, 2014, 23: 5428–5439
156 Caviedes J, Oberti F. A new sharpness metric based on local kurtosis, edge and energy information. Signal Process Image Commun, 2004, 19: 147–161
157 Ciancio A, da Costa A L N T, da Silva E A B, et al. No-reference blur assessment of digital pictures based on multifeature classifiers. IEEE Trans Image Process, 2011, 20: 64–75
158 Vu C T, Phan T D, Chandler D M. S3: a spectral and spatial measure of local perceived sharpness in natural images. IEEE Trans Image Process, 2012, 21: 934–945
159 Li L, Xia W, Lin W, et al. No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features. IEEE Trans Mult, 2017, 19: 1030–1040
160 Marziliano P, Dufaux F, Winkler S, et al. Perceptual blur and ringing metrics: application to JPEG2000. Signal Process Image Commun, 2004, 19: 163–172
161 Sheikh H R, Bovik A C, Cormack L. No-reference quality assessment using natural scene statistics: JPEG2000. IEEE Trans Image Process, 2005, 14: 1918–1927
162 Sazzad Z P, Kawayoke Y, Horita Y. Spatial features based no reference image quality assessment for JPEG2000. In: Proceedings of IEEE International Conference on Image Processing, 2007. 517
163 Sazzad Z M P, Kawayoke Y, Horita Y. No reference image quality assessment for JPEG2000 based on spatial features. Signal Process Image Commun, 2008, 23: 257–268
164 Zhang J, Le T. A new no-reference quality metric for JPEG2000 images. IEEE Trans Consumer Electron, 2010, 56: 743–750
165 Liu H, Klomp N, Heynderickx I. A no-reference metric for perceived ringing artifacts in images. IEEE Trans Circ Syst Video Technol, 2010, 20: 529–539
166 Liang L, Wang S, Chen J, et al. No-reference perceptual image quality metric using gradient profiles for JPEG2000. Signal Process Image Commun, 2010, 25: 502–516
167 Zhang J, Ong S H, Le T M. Kurtosis-based no-reference quality assessment of JPEG2000 images. Signal Process Image Commun, 2011, 26: 13–23
168 Moorthy A K, Bovik A C. A two-step framework for constructing blind image quality indices. IEEE Signal Process Lett, 2010, 17: 513–516
169 Moorthy A K, Bovik A C. Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans Image Process, 2011, 20: 3350–3364
170 Tang H, Joshi N, Kapoor A. Learning a blind measure of perceptual image quality. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011. 305–312
171 Gao X B, Gao F, Tao D C, et al. Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning. IEEE Trans Neural Netw Learn Syst, 2013, 24: 2013–2026
172 Zhang Y, Moorthy A K, Chandler D M, et al. C-DIIVINE: no-reference image quality assessment based on local magnitude and phase statistics of natural scenes. Signal Process Image Commun, 2014, 29: 725–747
173 Wang Q, Chu J, Xu L, et al. A new blind image quality framework based on natural color statistic. Neurocomputing, 2016, 173: 1798–1810
174 Saad M A, Bovik A C, Charrier C. A DCT statistics-based blind image quality index. IEEE Signal Process Lett, 2010, 17: 583–586
175 Saad M A, Bovik A C, Charrier C. Blind image quality assessment: a natural scene statistics approach in the DCT domain. IEEE Trans Image Process, 2012, 21: 3339–3352

176 Mittal A, Moorthy A K, Bovik A C. No-reference image quality assessment in the spatial domain. IEEE Trans Image Process, 2012, 21: 4695–4708
177 Mittal A, Soundararajan R, Bovik A C. Making a "completely blind" image quality analyzer. IEEE Signal Process Lett, 2013, 20: 209–212
178 Zhang L, Zhang L, Bovik A C. A feature-enriched completely blind image quality evaluator. IEEE Trans Image Process, 2015, 24: 2579–2591
179 Xue W, Mou X, Zhang L, et al. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. IEEE Trans Image Process, 2014, 23: 4850–4862
180 Lee D, Plataniotis K N. Toward a no-reference image quality assessment using statistics of perceptual color descriptors. IEEE Trans Image Process, 2016, 25: 3875–3889
181 Mittal A, Moorthy A K, Bovik A C. Making image quality assessment robust. In: Proceedings of the 46th Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2012. 1718–1722
182 Wu Q, Wang Z, Li H. A highly efficient method for blind image quality assessment. In: Proceedings of IEEE International Conference on Image Processing, 2015. 339–343
183 Zhang M, Muramatsu C, Zhou X, et al. Blind image quality assessment using the joint statistics of generalized local binary pattern. IEEE Signal Process Lett, 2015, 22: 207–210
184 Lu W, Zeng K, Tao D, et al. No-reference image quality assessment in contourlet domain. Neurocomputing, 2010, 73: 784–794
185 Shen J, Li Q, Erlebacher G. Hybrid no-reference natural image quality assessment of noisy, blurry, JPEG2000, and JPEG images. IEEE Trans Image Process, 2011, 20: 2089–2098
186 Zhang Y, Chandler D M. No-reference image quality assessment based on log-derivative statistics of natural scenes. J Electron Imag, 2013, 22: 043025
187 Ye P, Doermann D. No-reference image quality assessment based on visual codebook. In: Proceedings of IEEE International Conference on Image Processing, 2011. 3089–3092
188 Ye P, Doermann D. No-reference image quality assessment using visual codebooks. IEEE Trans Image Process, 2012, 21: 3129–3138
189 Ye P, Kumar J, Kang L, et al. Unsupervised feature learning framework for no-reference image quality assessment. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012. 1098–1105
190 Ye P, Kumar J, Kang L, et al. Real-time no-reference image quality assessment based on filter learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013. 987–994
191 Xue W, Zhang L, Mou X. Learning without human scores for blind image quality assessment. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013. 995–1002
192 Ye P, Kumar J, Doermann D. Beyond human opinion scores: blind image quality assessment based on synthetic scores. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014. 4241–4248
193 Xu J, Ye P, Li Q, et al. Blind image quality assessment based on high order statistics aggregation. IEEE Trans Image Process, 2016, 25: 4444–4457
194 Zhang P, Zhou W, Wu L, et al. SOM: semantic obviousness metric for image quality assessment. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015. 2394–2402
195 Mittal A, Muralidhar G S, Ghosh J, et al. Blind image quality assessment without human training using latent quality factors. IEEE Signal Process Lett, 2012, 19: 75–78
196 Jiang Q, Shao F, Jiang G, et al. Supervised dictionary learning for blind image quality assessment using quality-constraint sparse coding. J Visual Commun Image Represent, 2015, 33: 123–133
197 Xie X, Zhang Y, Wu J, et al. Bag-of-words feature representation for blind image quality assessment with local quantized pattern. Neurocomputing, 2017, 266: 176–187
198 Jiang Q, Shao F, Lin W, et al. Optimizing multistage discriminative dictionaries for blind image quality assessment. IEEE Trans Mult, 2018, 20: 2035–2048
199 He L, Tao D, Li X, et al. Sparse representation for blind image quality assessment. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012. 1146–1153
200 Zhang C, Pan J, Chen S, et al. No reference image quality assessment using sparse feature representation in two dimensions spatial correlation. Neurocomputing, 2016, 173: 462–470
201 Wu Q, Li H, Meng F, et al. Blind image quality assessment based on multichannel feature fusion and label transfer. IEEE Trans Circ Syst Video Technol, 2016, 26: 425–440
202 Wu Q, Li H, Meng F, et al. No reference image quality assessment metric via multi-domain structural information and piecewise regression. J Visual Commun Image Represent, 2015, 32: 205–216
203 Wu Q, Li H, Ngan K N, et al. Blind image quality assessment using local consistency aware retriever and uncertainty aware evaluator. IEEE Trans Circ Syst Video Tech, 2018, 28: 2078–2089
204 Fang R, Al-Bayaty R, Wu D. BNB method for no-reference image quality assessment. IEEE Trans Circ Syst Video Technol, 2017, 27: 1381–1391
205 Gao F, Tao D, Gao X, et al. Learning to rank for blind image quality assessment. IEEE Trans Neural Netw Learn Syst, 2015, 26: 2275–2290
206 Ma K, Liu W, Liu T, et al. dipIQ: blind image quality assessment by learning-to-rank discriminable image pairs. IEEE Trans Image Process, 2017, 26: 3951–3964
207 Burges C, Shaked T, Renshaw E, et al. Learning to rank using gradient descent. In: Proceedings of International Conference on Machine Learning, 2005. 89–96

208 Xu L, Li J, Lin W, et al. Multi-task rank learning for image quality assessment. IEEE Trans Circ Syst Video Technol, 2017, 27: 1833–1843
209 Pang Y, Zhou B, Nie F. Simultaneously learning neighborship and projection matrix for supervised dimensionality reduction. IEEE Trans Neural Netw Learn Syst, 2019, 30: 2779–2793
210 Pang Y, Cao J, Li X. Cascade learning by optimally partitioning. IEEE Trans Cyber, 2016, 47: 4148–4161
211 Han Z Y, Wu H B, Wei B Z, et al. Recursive narrative alignment for movie narrating. Sci China Inf Sci, 2020, 63: 174101
212 Zhang W T, Jiang J W, Shao Y X, et al. Snapshot boosting: a fast ensemble framework for deep neural networks. Sci China Inf Sci, 2020, 63: 112102
213 Habimana O, Li Y H, Li R H, et al. Sentiment analysis using deep learning approaches: an overview. Sci China Inf Sci, 2020, 63: 111102
214 Chen S T, Jian Z Q, Huang Y H, et al. Autonomous driving: cognitive construction and situation understanding. Sci China Inf Sci, 2019, 62: 081101
215 Gu K, Zhai G, Yang X, et al. Deep learning network for blind image quality assessment. In: Proceedings of IEEE International Conference on Image Processing, 2014. 511–515
216 Li Y, Po L M, Xu X, et al. No-reference image quality assessment with shearlet transform and deep neural networks. Neurocomputing, 2015, 154: 94–109
217 Lv Y, Jiang G, Yu M, et al. Difference of gaussian statistical features based blind image quality assessment: a deep learning approach. In: Proceedings of IEEE International Conference on Image Processing, 2015. 2344–2348
218 Li C F, Bovik A C, Wu X J. Blind image quality assessment using a general regression neural network. IEEE Trans Neural Netw, 2011, 22: 793–799
219 Specht D F. A general regression neural network. IEEE Trans Neural Netw, 1991, 2: 568–576
220 Tang H, Joshi N, Kapoor A. Blind image quality assessment using semi-supervised rectifier networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014. 2877–2884
221 Hinton G E. Reducing the dimensionality of data with neural networks. Science, 2006, 313: 504–507
222 Hou W L, Gao X B, Tao D C, et al. Blind image quality assessment via deep learning. IEEE Trans Neural Netw Learn Syst, 2015, 26: 1275–1286
223 Kang L, Ye P, Li Y, et al. Convolutional neural networks for no-reference image quality assessment. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014. 1733–1740
224 Bosse S, Maniry D, Wiegand T, et al. A deep neural network for image quality assessment. In: Proceedings of IEEE International Conference on Image Processing, 2016. 3773–3777
225 Kang L, Ye P, Li Y, et al. Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks. In: Proceedings of IEEE International Conference on Image Processing, 2015. 2791–2795
226 Kim J, Lee S. Fully deep blind image quality predictor. IEEE J Sel Top Signal Process, 2017, 11: 206–220
227 Gu J, Meng G, Redi J A, et al. Blind image quality assessment via vector regression and object oriented pooling. IEEE Trans Mult, 2018, 20: 1140–1153
228 Kim J, Nguyen A D, Lee S. Deep CNN-based blind image quality predictor. IEEE Trans Neural Netw Learn Syst, 2019, 30: 11–24
229 Pan D, Shi P, Hou M, et al. Blind predicting similar quality map for image quality assessment. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018. 6373–6382
230 Ma K, Liu W, Zhang K, et al. End-to-end blind image quality assessment using deep neural networks. IEEE Trans Image Process, 2018, 27: 1202–1213
231 Lin K Y, Wang G. Hallucinated-iqa: no-reference image quality assessment via adversarial learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018. 732–741
232 Liu X, van de Weijer J, Bagdanov A D. Rankiqa: learning from rankings for no-reference image quality assessment. In: Proceedings of IEEE International Conference on Computer Vision, 2017. 1040–1049
233 Talebi H, Milanfar P. NIMA: neural image assessment. IEEE Trans Image Process, 2018, 27: 3998–4011
234 Guan J, Yi S, Zeng X, et al. Visual importance and distortion guided deep image quality assessment framework. IEEE Trans Mult, 2017, 19: 2505–2520
235 Zhai G, Wu X, Yang X, et al. A psychovisual quality metric in free-energy principle. IEEE Trans Image Process, 2012, 21: 41–52
236 Zhai G, Min X, Liu N. Free-energy principle inspired visual quality assessment: an overview. Digital Signal Process, 2019, 91: 11–20
237 Gu K, Zhai G, Yang X, et al. Using free energy principle for blind image quality assessment. IEEE Trans Mult, 2015, 17: 50–63
238 Li Q, Lin W, Xu J, et al. Blind image quality assessment using statistical structural and luminance features. IEEE Trans Mult, 2016, 18: 2457–2469
239 Li Q, Lin W, Fang Y. BSD: blind image quality assessment based on structural degradation. Neurocomputing, 2017, 236: 93–103
240 Min X, Gu K, Zhai G, et al. Blind quality assessment based on pseudo-reference image. IEEE Trans Mult, 2018, 20: 2049–2062
241 Min X, Zhai G, Gu K, et al. Blind image quality estimation via distortion aggravation. IEEE Trans Broadcast, 2018, 64: 508–517
242 Wu Q, Li H, Wang Z, et al. Blind image quality assessment based on rank-order regularized regression. IEEE Trans

Mult, 2017, 19: 2490–2504
243 Saha A, Wu Q M J. Utilizing image scales towards totally training free blind image quality assessment. IEEE Trans Image Process, 2015, 24: 1879–1892
244 Liu T J, Liu K H. No-reference image quality assessment by wide-perceptual-domain scorer ensemble method. IEEE Trans Image Process, 2018, 27: 1138–1151
245 Freitas P G, Akamine W Y L, Farias M C Q. No-reference image quality assessment using orthogonal color planes patterns. IEEE Trans Mult, 2018, 20: 3353–3360
246 Lambooij M, IJsselsteijn W, Bouwhuis D G, et al. Evaluation of stereoscopic images: beyond 2D quality. IEEE Trans Broadcast, 2011, 57: 432–444
247 Wang J, Wang S, Ma K, et al. Perceptual depth quality in distorted stereoscopic images. IEEE Trans Image Process, 2017, 26: 1202–1215
248 Yun N, Feng Z, Yang J, et al. The objective quality assessment of stereo image. Neurocomputing, 2013, 120: 121–129
249 You J, Xing L, Perkis A, et al. Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. In: Proceedings of International Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2010
250 Akhter R, Sazzad Z M P, Horita Y, et al. No-reference stereoscopic image quality assessment. In: Proceedings of SPIE, 2010. 7524
251 Hewage C T E R, Martini M G. Reduced-reference quality metric for 3D depth map transmission. In: Proceedings of 3DTV-Conference: the True Vision - Capture, Transmission and Display of 3D Video, 2010. 1–4
252 Maalouf A, Larabi M C. CYCLOP: a stereo color image quality assessment metric. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2011. 1161–1164
253 Shao F, Tian W, Lin W, et al. Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions. IEEE Trans Image Process, 2016, 25: 2059–2074
254 Yang J, Zhao Y, Zhu Y, et al. Blind assessment for stereo images considering binocular characteristics and deep perception map based on deep belief network. Inf Sci, 2019, 474: 1–17
255 Stidwill D, Fletcher R. Normal Binocular Vision: Theory, Investigation and Practical Aspects. Hoboken: John Wiley & Sons, 2010
256 Li Z, Atick J J. Efficient stereo coding in the multiscale representation. Netw Comput Neural Syst, 1994, 5: 157–174
257 Moorthy A K, Su C C, Mittal A, et al. Subjective evaluation of stereoscopic image quality. Signal Process Image Commun, 2013, 28: 870–883
258 Gorley P, Holliman N. Stereoscopic image quality metrics and compression. In: Proceedings of SPIE, 2008. 680305
259 Yasakethu S L P, Hewage C T E R, Fernando W A C, et al. Quality analysis for 3D video using 2D video quality models. IEEE Trans Consumer Electron, 2008, 54: 1969–1976
260 Meegan D V, Stelmach L B, Tam W J. Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery. J Exp Psychol Appl, 2001, 7: 143–153
261 Fang Y, Yan J, Liu X, et al. Stereoscopic image quality assessment by deep convolutional neural network. J Visual Commun Image Represent, 2019, 58: 400–406
262 Zhou W, Chen Z, Li W. Dual-stream interactive networks for no-reference stereoscopic image quality assessment. IEEE Trans Image Process, 2019, 28: 3946–3958
263 Kumano H, Tanabe S, Fujita I. Spatial frequency integration for binocular correspondence in macaque area V4. J Neuro Physiol, 2008, 99: 402–408
264 Lin Y H, Wu J L. Quality assessment of stereoscopic 3D image compression by binocular integration behaviors. IEEE Trans Image Process, 2014, 23: 1527–1542
265 Jiang G, Xu H, Yu M, et al. Stereoscopic image quality assessment by learning non-negative matrix factorization-based color visual characteristics and considering binocular interactions. J Visual Commun Image Represent, 2017, 46: 269–279
266 Kingdom F A A. Binocular vision: the eyes add and subtract. Curr Biol, 2012, 22: 22–24
267 Yang J, Liu Y, Gao Z, et al. A perceptual stereoscopic image quality assessment model accounting for binocular combination behavior. J Visual Commun Image Represent, 2015, 31: 138–145
268 Lin C, Chen Z, Liao N. Full-reference quality assessment for stereoscopic images based on binocular vision model. In: Proceedings of IEEE International Conference on Visual Communications and Image Processing, 2016. 1–4
269 Qian N, Mikaelian S. Relationship between phase and energy methods for disparity computation. Neural Comput, 2000, 12: 279–292
270 Field D J. Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am Opt Image Sci, 1987, 4: 2379
271 Lin Y, Yang J, Lu W, et al. Quality index for stereoscopic images by jointly evaluating cyclopean amplitude and cyclopean phase. IEEE J Sel Top Signal Process, 2017, 11: 89–101
272 Kruger N, Janssen P, Kalkan S, et al. Deep hierarchies in the primate visual cortex: what can we learn for computer vision? IEEE Trans Pattern Anal Mach Intell, 2013, 35: 1847–1871
273 Barlow H B. Foundations of cyclopean perception. Behav Sci, 1974, 17: 310–312
274 Grossberg S, Kelly F. Neural dynamics of binocular brightness perception. Vision Res, 1999, 39: 3796–3816
275 Liu Y, Yang J, Meng Q, et al. Stereoscopic image quality assessment method based on binocular combination saliency model. Signal Process, 2016, 125: 237–248
276 Zhou W, Jiang G, Yu M, et al. PMFS: a perceptual modulated feature similarity metric for stereoscopic image quality

assessment. IEEE Signal Process Lett, 2014, 21: 1003–1006
277 Li S, Han X, Chang Y. Adaptive cyclopean image-based stereoscopic image-quality assessment using ensemble learning. IEEE Trans Mult, 2019, 21: 2616–2624
278 Md S K, Appina B, Channappayya S S. Full-reference stereo image quality assessment using natural stereo scene statistics. IEEE Signal Process Lett, 2015, 22: 1985–1989
279 Ko H, Song R, Jay Kuo C C. A ParaBoost stereoscopic image quality assessment (PBSIQA) system. J Visual Commun Image Represent, 2017, 45: 156–169
280 Wang X, Liu Q, Wang R, et al. Natural image statistics based 3D reduced reference image quality assessment in contourlet domain. Neurocomputing, 2015, 151: 683–691
281 Su C C, Cormack L K, Bovik A C. Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation. IEEE Trans Image Process, 2015, 24: 1685–1699
282 Ma L, Wang X, Liu Q, et al. Reorganized DCT-based image representation for reduced reference stereoscopic image quality assessment. Neurocomputing, 2016, 215: 21–31
283 Zhou W, Yu L. Binocular responses for no-reference 3D image quality assessment. IEEE Trans Mult, 2016, 18: 1077–1084
284 Zhou W, Qiu W, Wu M W. Utilizing dictionary learning and machine learning for blind quality assessment of 3-D images. IEEE Trans Broadcast, 2017, 63: 404–415
285 Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Res, 1997, 37: 3311–3325
286 Elad M. Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Berlin: Springer, 2010
287 Shao F, Lin W S, Wang S S, et al. Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook. IEEE Trans Broadcast, 2015, 61: 154–165
288 Shao F, Li K, Lin W, et al. Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties. IEEE Trans Image Process, 2015, 24: 2971–2983
289 Qi F, Zhao D, Gao W. Reduced reference stereoscopic image quality assessment based on binocular perceptual information. IEEE Trans Mult, 2015, 17: 2338–2344
290 Shao F, Li K, Lin W, et al. Learning blind quality evaluator for stereoscopic images using joint sparse representation. IEEE Trans Mult, 2016, 18: 2104–2114
291 Shao F, Li K, Lin W, et al. Using binocular feature combination for blind quality assessment of stereoscopic images. IEEE Signal Process Lett, 2015, 22: 1548–1551
292 Shao F, Lin W, Wang S, et al. Learning receptive fields and quality lookups for blind quality assessment of stereoscopic images. IEEE Trans Cybern, 2016, 46: 730–743
293 Shao F, Zhang Z, Jiang Q, et al. Toward domain transfer for no-reference quality prediction of asymmetrically distorted stereoscopic images. IEEE Trans Circ Syst Video Technol, 2018, 28: 573–585
294 Shao F, Tian W, Lin W, et al. Learning sparse representation for no-reference quality assessment of multiply distorted stereoscopic images. IEEE Trans Mult, 2017, 19: 1821–1836
295 Vu C T, Larson E C, Chandler D M. Visual fixation patterns when judging image quality: effects of distortion type, amount, and subject experience. In: Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation, 2008. 73–76
296 Liu H T, Heynderickx I. Visual attention in objective image quality assessment: based on eye-tracking data. IEEE Trans Circ Syst Video Technol, 2011, 21: 971–982
297 Liu H T, Engelke U, Wang J, et al. How does image content affect the added value of visual attention in objective image quality assessment? IEEE Signal Process Lett, 2013, 20: 355–358
298 Wang Q, Xu L, Chen Q, et al. Import of distortion on saliency applied to image quality assessment. In: Proceedings of IEEE International Conference on Image Processing, 2014. 1165–1169
299 Rai Y, Callet P L, Guillotel P. Which saliency weighting for omni directional image quality assessment? In: Proceedings of IEEE International Conference on Quality of Multimedia Experience, 2017. 1–6
300 Zhang W, Liu H. Toward a reliable collection of eye-tracking data for image quality research: challenges, solutions, and applications. IEEE Trans Image Process, 2017, 26: 2424–2437
301 Ma Q, Zhang L. Image quality assessment with visual attention. In: Proceedings of IEEE International Conference on Pattern Recognition, 2008. 1–4
302 Zhang W, Talens-Noguera J V, Liu H. The quest for the integration of visual saliency models in objective image quality assessment: a distraction power compensated combination strategy. In: Proceedings of IEEE International Conference on Image Processing, 2015. 1250–1254
303 Zhang W, Borji A, Wang Z, et al. The application of visual saliency models in objective image quality assessment: a statistical evaluation. IEEE Trans Neural Netw Learn Syst, 2016, 27: 1266–1278
304 Wen Y, Li Y, Zhang X, et al. A weighted full-reference image quality assessment based on visual saliency. J Visual Commun Image Represent, 2017, 43: 119–126
305 Xia Y, Liu Z, Yan Y, et al. Media quality assessment by perceptual gaze-shift patterns discovery. IEEE Trans Mult, 2017, 19: 1811–1820
306 Zhang W, Martin R R, Liu H. A saliency dispersion measure for improving saliency-based image quality metrics. IEEE Trans Circ Syst Video Technol, 2018, 28: 1462–1466
307 Mittal A, Moorthy A K, Bovik A C, et al. Automatic prediction of saliency on JPEG distorted images. In: Proceedings

of IEEE International Workshop on Quality of Multimedia Experience, 2011. 195–200
308 Winterlich A, Zlokolica V, Denny P, et al. A saliency weighted no-reference perceptual blur metric for the automotive environment. In: Proceedings of IEEE International Workshop on Quality of Multimedia Experience, 2013. 206–211
309 Harel J, Koch C, Perona P. Graph-based visual saliency. In: Proceedings of the 20th Annual Conference on Neural Information Processing Systems, 2006. 545–552
310 Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Machine Intell, 1998, 20: 1254–1259
311 Nasrinpour H R, Bruce N D. Saliency weighted quality assessment of tone-mapped images. In: Proceedings of IEEE International Conference on Image Processing, 2015. 4947–4951
312 Bruce N T J. Attention based on information maximization. J Vision, 2007, 7: 950–950
313 Kundu D, Evans B L. Visual attention guided quality assessment of tone-mapped images using scene statistics. In: Proceedings of IEEE International Conference on Image Processing, 2016. 96–100
314 Min X, Zhai G, Gao Z, et al. Influence of compression artifacts on visual attention. In: Proceedings of IEEE International Conference on Multimedia and Expo, 2014. 1–6
315 Che Z, Borji A, Zhai G, et al. How is gaze influenced by image transformations? Dataset and model. IEEE Trans Image Process, 2020, 29: 2287–2300
316 Che Z, Zhai G, Min X. Influence of spatial resolution on state-of-the-art saliency models. In: Proceedings of Pacific Rim Conference on Multimedia, 2015. 74–83
317 Coutrot A, Guyader N. How saliency, faces, and sound influence gaze in dynamic social scenes. J Vision, 2014, 14: 5
318 Min X, Zhai G, Gu K, et al. Fixation prediction through multimodal analysis. ACM Trans Mult Comput Commun Appl, 2017, 13: 6
319 Min X, Zhai G, Gao Z, et al. Sound influences visual attention discriminately in videos. In: Proceedings of IEEE International Workshop on Quality of Multimedia Experience, 2014. 153–158
320 Min X, Zhai G, Zhou J, et al. A multimodal saliency model for videos with high audio-visual correspondence. IEEE Trans Image Process, 2020, 29: 3805–3819
321 Min X, Zhai G, Gu K, et al. Visual attention analysis and prediction on human faces. Inf Sci, 2017, 420: 417–430
322 Duan H, Min X, Fang Y, et al. Visual attention analysis and prediction on human faces for children with autism spectrum disorder. ACM Trans Mult Comput Commun Appl, 2019, 15: 90
323 Wang S, Jiang M, Duchesne X M, et al. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. Neuron, 2015, 88: 604–616
324 Duan H, Zhai G, Min X, et al. A dataset of eye movements for the children with autism spectrum disorder. In: Proceedings of the 10th ACM Multimedia Systems Conference, 2019. 255–260
325 Duan H, Zhai G, Min X, et al. Learning to predict where the children with asd look. In: Proceedings of IEEE International Conference on Image Processing, 2018, 704–708
326 Hou W, Gao X. Be natural: a saliency-guided deep framework for image quality. In: Proceedings of IEEE International Conference on Multimedia and Expo, 2014. 1–6
327 Zhang L, Shen Y, Li H. VSI: a visual saliency-induced index for perceptual image quality assessment. IEEE Trans Image Process, 2014, 23: 4270–4281
328 Zhang L, Gu Z, Li H. SDSP: a novel saliency detection method by combining simple priors. In: Proceedings of IEEE International Conference on Image Processing, 2013. 171–175
329 Zhang W, Liu H. Learning picture quality from visual distraction: Psychophysical studies and computational models. Neurocomputing, 2017, 247: 183–191
330 Yang H, Fang Y, Lin W. Perceptual quality assessment of screen content images. IEEE Trans Image Process, 2015, 24: 4408–4421
331 Gu K, Wang S, Zhai G, et al. Screen image quality assessment incorporating structural degradation measurement. In: Proceedings of IEEE International Symposium on Circuits and Systems, 2015. 125–128
332 Wang S, Gu K, Zeng K, et al. Perceptual screen content image quality assessment and compression. In: Proceedings of IEEE International Conference on Image Processing, 2015. 1434–1438
333 Wang S, Gu K, Zeng K, et al. Objective quality assessment and perceptual compression of screen content images. IEEE Comput Grap Appl, 2018, 38: 47–58
334 Ni Z, Ma L, Zeng H, et al. Gradient direction for screen content image quality assessment. IEEE Signal Process Lett, 2016, 23: 1394–1398
335 Ni Z, Ma L, Zeng H, et al. Screen content image quality assessment using edge model. In: Proceedings of IEEE International Conference on Image Processing, 2016. 81–85
336 Ni Z, Zeng H, Ma L, et al. A Gabor feature-based quality assessment model for the screen content images. IEEE Trans Image Process, 2018, 27: 4516–4528
337 Fu Y, Zeng H, Ma L, et al. Screen content image quality assessment using multi-scale difference of Gaussian. IEEE Trans Circ Syst Video Technol, 2018, 28: 2428–2432
338 Gu K, Wang S, Yang H, et al. Saliency-guided quality assessment of screen content images. IEEE Trans Mult, 2016, 18: 1098–1110
339 Gu K, Qiao J, Min X, et al. Evaluating quality of screen content images via structural variation analysis. IEEE Trans Visual Comput Graph, 2017, 24: 2689–2701
340 Fang Y, Yan J, Liu J, et al. Objective quality assessment of screen content images by uncertainty weighting. IEEE Trans Image Process, 2017, 26: 2016–2027

341 Zhang Y, Chandler D M, Mou X. Quality assessment of screen content images via convolutional-neural-network-based synthetic/natural segmentation. IEEE Trans Image Process, 2018, 27: 5113–5128
342 Wang S, Gu K, Zhang X, et al. Subjective and objective quality assessment of compressed screen content images. IEEE J Emerg Sel Top Circ Syst, 2016, 6: 532–543
343 Wang S, Gu K, Zhang X, et al. Reduced-reference quality assessment of screen content images. IEEE Trans Circ Syst Video Tech, 2018, 28: 1–14
344 Jakhetiya V, Gu K, Lin W, et al. A prediction backed model for quality assessment of screen content and 3-D synthesized images. IEEE Trans Ind Inf, 2018, 14: 652–660
345 Gu K, Zhai G, Lin W, et al. Learning a blind quality evaluation engine of screen content images. Neurocomputing, 2016, 196: 140–149
346 Zuo L, Wang H, Fu J. Screen content image quality assessment via convolutional neural network. In: Proceedings of IEEE International Conference on Image Processing, 2016. 2082–2086
347 Gu K, Zhou J, Qiao J F, et al. No-reference quality assessment of screen content pictures. IEEE Trans Image Process, 2017, 26: 4005–4018
348 Shao F, Gao Y, Li F, et al. Toward a blind quality predictor for screen content images. IEEE Trans Syst Man Cyber Syst, 2018, 48: 1521–1530
349 Fang Y, Yan J, Li L, et al. No reference quality assessment for screen content images with both local and global feature representation. IEEE Trans Image Process, 2018, 27: 1600–1610
350 Zhou W, Yu L, Zhou Y, et al. Local and global feature learning for blind quality evaluation of screen content and natural scene images. IEEE Trans Image Process, 2018, 27: 2086–2095
351 Ma K, Yeganeh H, Zeng K, et al. High dynamic range image compression by optimizing tone mapped image quality index. IEEE Trans Image Process, 2015, 24: 3086–3097
352 Gu K, Wang S, Zhai G, et al. Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure. IEEE Trans Mult, 2016, 18: 432–443
353 Gu K, Zhai G, Liu M, et al. Details preservation inspired blind quality metric of tone mapping methods. In: Proceedings of IEEE International Symposium on Circuits and Systems, 2014. 518–521
354 Nafchi H Z, Shahkolaei A, Moghaddam R F, et al. FSITM: a feature similarity index for tone-mapped images. IEEE Signal Process Lett, 2015, 22: 1026–1029
355 Kundu D, Ghadiyaram D, Bovik A C, et al. No-reference quality assessment of tone-mapped HDR pictures. IEEE Trans Image Process, 2017, 26: 2957–2971
356 Hadizadeh H, Bajic I V. Full-reference objective quality assessment of tone-mapped images. IEEE Trans Mult, 2018, 20: 392–404
357 Yue G, Hou C, Gu K, et al. Biologically inspired blind quality assessment of tone-mapped images. IEEE Trans Indust Electron, 2018, 65: 2525–2536
358 Yue G, Hou C, Zhou T. Blind quality assessment of tone-mapped images considering colorfulness, naturalness and structure. IEEE Trans Indust Electron, 2019, 66: 3784–3793
359 Xydeas C S, Petrovic V S. Objective pixel-level image fusion performance measure. In: Proceedings of International Society for Optics and Photonics, 2000. 89–98
360 Qu G, Zhang D, Yan P. Information measure for performance of image fusion. Electron Lett, 2002, 38: 313–315
361 Piella G, Heijmans H. A new quality metric for image fusion. In: Proceedings of IEEE International Conference on Image Processing, 2003. 173
362 Cvejic N, Canagarajah C N, Bull D R. Image fusion metric based on mutual information and Tsallis entropy. Electron Lett, 2006, 42: 626–627
363 Chen H, Varshney P K. A human perception inspired quality metric for image fusion based on regional information. Inf Fusion, 2007, 8: 193–207
364 Zheng Y, Essock E A, Hansen B C, et al. A new metric based on extended spatial frequency and its application to DWT based fusion algorithms. Inf Fusion, 2007, 8: 177–192
365 Wang P W, Liu B. A novel image fusion metric based on multi-scale analysis. In: Proceedings of IEEE International Conference on Signal Processing, 2008. 965–968
366 Hossny M, Nahavandi S, Creighton D. Comments on 'Information measure for performance of image fusion'. Electron Lett, 2008, 44: 1066–1067
367 Chen Y, Blum R S. A new automated quality assessment algorithm for image fusion. Image Vision Comput, 2009, 27: 1421–1432
368 Hassen R, Wang Z, Salama M M A. Objective quality assessment for multiexposure multifocus image fusion. IEEE Trans Image Process, 2015, 24: 2712–2724
369 Karimi M, Samavi S, Karimi N, et al. Quality assessment of retargeted images by salient region deformity analysis. J Visual Commun Image Represent, 2017, 43: 108–118
370 Ma L, Xu L, Zhang Y, et al. No-reference retargeted image quality assessment based on pairwise rank learning. IEEE Trans Mult, 2016, 18: 2228–2237
371 Zhang Y, Fang Y, Lin W, et al. Backward registration-based aspect ratio similarity for image retargeting quality assessment. IEEE Trans Image Process, 2016, 25: 4286–4297
372 Fang Y, Zeng K, Wang Z, et al. Objective quality assessment for image retargeting based on structural similarity. IEEE J Emerg Sel Top Circ Syst, 2014, 4: 95–105
373 Hsu C C, Lin C W, Fang Y, et al. Objective quality assessment for image retargeting based on perceptual geometric

distortion and information loss. IEEE J Sel Top Signal Process, 2014, 8: 377–389
374 Chen Z, Lin J, Liao N, et al. Full reference quality assessment for image retargeting based on natural scene statistics modeling and bi-directional saliency similarity. IEEE Trans Image Process, 2017, 26: 5138–5148
375 Zhang Y, Lin W, Li Q, et al. Multiple-level feature-based measure for retargeted image quality. IEEE Trans Image Process, 2018, 27: 451–463
376 Zhang Y, Ngan K N, Ma L, et al. Objective quality assessment of image retargeting by incorporating fidelity measures and inconsistency detection. IEEE Trans Image Process, 2017, 26: 5980–5993
377 Liang Y, Liu Y J, Gutierrez D. Objective quality prediction of image retargeting algorithms. IEEE Trans Visual Comput Graph, 2017, 23: 1099–1110
378 Zhang F, Roysam B. Blind quality metric for multidistortion images based on cartoon and texture decomposition. IEEE Signal Process Lett, 2016, 23: 1265–1269
379 Lu Y, Xie F, Liu T, et al. No reference quality assessment for multiply-distorted images based on an improved bag-of-words model. IEEE Signal Process Lett, 2015, 22: 1811–1815
380 Hadizadeh H, Bajic I V. Color Gaussian jet features for no-reference quality assessment of multiply-distorted images. IEEE Signal Process Lett, 2016, 23: 1717–1721
381 Li Q, Lin W, Fang Y. No-reference quality assessment for multiply-distorted images in gradient domain. IEEE Signal Process Lett, 2016, 23: 541–545
382 Zhang Y, Chandler D M. Opinion-unaware blind quality assessment of multiply and singly distorted images via distortion parameter estimation. IEEE Trans Image Process, 2018, 27: 5433–5448
383 Brooks A C, Zhao X N, Pappas T N. Structural similarity quality metrics in a coding context: exploring the space of realistic distortions. IEEE Trans Image Process, 2008, 17: 1261–1273
384 Yang L, Du H, Xu J, et al. Blind image quality assessment on authentically distorted images with perceptual features. In: Proceedings of IEEE International Conference on Image Processing, 2016. 2042–2046
385 Ghadiyaram D, Bovik A C. Scene statistics of authentically distorted images in perceptually relevant color spaces for blind image quality assessment. In: Proceedings of IEEE International Conference on Image Processing, 2015. 3851–3855
386 Ghadiyaram D, Bovik A C. Perceptual quality prediction on authentically distorted images using a bag of features approach. J Vision, 2017, 17: 32–32
387 Liu Y, Gu K, Wang S, et al. Blind quality assessment of camera images based on low-level and high-level statistical features. IEEE Trans Mult, 2019, 21: 135–146
388 Sinno Z, Bovik A C. Large-scale study of perceptual video quality. IEEE Trans Image Process, 2018, 28: 612–627
389 Battisti F, Bosc E, Carli M, et al. Objective image quality assessment of 3D synthesized views. Signal Process Image Commun, 2015, 30: 78–88
390 Sandić-Stanković D, Kukolj D, Le Callet P. Multi-scale synthesized view assessment based on morphological pyramids. J Electr Eng, 2016, 67: 3–11
391 Sandić-Stanković D, Kukolj D, Callet P. DIBR-synthesized image quality assessment based on morphological multi-scale approach. EURASIP J Image Video Process, 2016, 2017: 4
392 Li L, Zhou Y, Gu K, et al. Quality assessment of DIBR-synthesized images by measuring local geometric distortions and global sharpness. IEEE Trans Mult, 2018, 20: 914–926
393 Gu K, Jakhetiya V, Qiao J F, et al. Model-based referenceless quality metric of 3D synthesized images using local image description. IEEE Trans Image Process, 2018, 27: 394–405
394 Tian S, Zhang L, Morin L, et al. NIQSV+: a no-reference synthesized view quality assessment metric. IEEE Trans Image Process, 2018, 27: 1652–1664
395 Zhou Y, Li L, Wang S, et al. No-reference quality assessment of DIBR-synthesized videos by measuring temporal flickering. J Visual Commun Image Represent, 2018, 55: 30–39
396 Ling S, Li J, Che Z, et al. Quality assessment of free-viewpoint videos by quantifying the elastic changes of multi-scale motion trajectories. 2019. ArXiv: 190312107
397 Li B, Ren W, Fu D, et al. Benchmarking single-image dehazing and beyond. IEEE Trans Image Process, 2018, 28: 492–505
398 Hautiere N, Tarel J P, Aubert D, et al. Blind contrast enhancement assessment by gradient ratioing at visible edges. Image Anal Stereol, 2008, 27: 87–95
399 Duan H, Zhai G, Yang X, et al. IVQAD 2017: an immersive video quality assessment database. In: Proceedings of International Conference on Systems, Signals and Image Processing, 2017. 1–5
400 Xu M, Li C, Chen Z, et al. Assessing visual quality of omnidirectional videos. IEEE Trans Circ Syst Video Technol, 2019, 29: 3516–3530
401 Duan H, Zhai G, Min X, et al. Assessment of visually induced motion sickness in immersive videos. In: Advances in Multimedia Information Processing—PCM 2017. Berlin: Springer, 2017. 662–672
402 Yu M, Lakshman H, Girod B. A framework to evaluate omnidirectional video coding schemes. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality, 2015. 31–36
403 Yule S, Lu A, Lu Y. WS-PSNR for 360 video objective quality evaluation. MPEG Joint Video Exploration Team, 2016. 116
404 Zakharchenko V, Choi K P, Park J H. Quality metric for spherical panoramic video. In: Proceedings of Optics and Photonics for Information Processing X. 2016. 99700C
405 Huang W, Ding L, Zhai G, et al. Utility-oriented resource allocation for 360-degree video transmission over hetero-

geneous networks. Digital Signal Process, 2019, 84: 1–14
406 Gutiérrez J, David E, Rai Y, et al. Toolbox and dataset for the development of saliency and scanpath models for omnidirectional/360° still images. Signal Process Image Commun, 2018, 69: 35–42
407 Zhu Y, Zhai G, Min X, et al. The prediction of saliency map for head and eye movements in 360 degree images. IEEE Trans Mult, 2019. doi: 10.1109/TMM.2019.2957986
408 Zhu Y, Zhai G, Min X. The prediction of head and eye movement for 360 degree images. Signal Process Image Commun, 2018, 69: 15–25
409 Fang Y, Zhang X, Imamoglu N. A novel superpixel-based saliency detection model for 360-degree images. Signal Process Image Commun, 2018, 69: 1–7
410 Ling J, Zhang K, Zhang Y, et al. A saliency prediction model on 360 degree images using color dictionary based sparse representation. Signal Process Image Commun, 2018, 69: 60–68
411 Monroy R, Lutz S, Chalasani T, et al. SalNet360: saliency maps for omni-directional images with CNN. Signal Process Image Commun, 2018, 69: 26–34
412 Lebreton P, Raake A. GBVS360, BMS360, ProSal: extending existing saliency prediction models from 2D to omnidirectional images. Signal Process Image Commun, 2018, 69: 69–78
413 Li C, Xu M, Zhang S, et al. State-of-the-art in 360° video/image processing: perception, assessment and compression. 2019. ArXiv: 190500161
414 Chikkerur S, Sundaram V, Reisslein M, et al. Objective video quality assessment methods: a classification, review, and performance comparison. IEEE Trans Broadcast, 2011, 57: 165–182
415 Seshadrinathan K, Soundararajan R, Bovik A C, et al. Study of subjective and objective quality assessment of video. IEEE Trans Image Process, 2010, 19: 1427–1441
416 Pinson M H, Wolf S. A new standardized method for objectively measuring video quality. IEEE Trans Brocast, 2004, 50: 312–322
417 Zhai G T, Cai J F, Lin W S, et al. Cross-dimensional perceptual quality assessment for low bit-rate videos. IEEE Trans Mult, 2008, 10: 1316–1324
418 Zhai G T, Cai J F, Lin W S, et al. Three dimensional scalable video adaptation via user-end perceptual quality assessment. IEEE Trans Broadcast, 2008, 54: 719–727
419 Zhao T, Liu Q, Chen C W. QoE in video transmission: a user experience-driven strategy. IEEE Commun Surv Tutor, 2016, 19: 285–302
420 Bampis C G, Li Z, Moorthy A K, et al. Study of temporal effects on subjective video quality of experience. IEEE Trans Image Process, 2017, 26: 5217–5231
421 Bampis C G, Li Z, Katsavounidis I, et al. Towards perceptually optimized end-to-end adaptive video streaming. 2018. ArXiv: 180803898
422 Ghadiyaram D, Pan J, Bovik A C. A subjective and objective study of stalling events in mobile streaming videos. IEEE Trans Circ Syst Video Technol, 2017, 29: 183–197
423 Duanmu Z, Ma K, Wang Z. Quality-of-experience for adaptive streaming videos: an expectation confirmation theory motivated approach. IEEE Trans Image Process, 2018, 27: 6135–6146
424 Duanmu Z, Rehman A, Wang Z. A quality-of-experience database for adaptive video streaming. IEEE Trans Broadcast, 2018, 64: 474–487
425 Duanmu Z, Zeng K, Ma K, et al. A quality-of-experience index for streaming video. IEEE J Sel Top Signal Process, 2016, 11: 154–166
426 Pinson M H, Janowski L, Pepion R, et al. The influence of subjects and environment on audiovisual subjective tests: an international study. IEEE J Sel Top Signal Process, 2012, 6: 640–651
427 Akhtar Z, Falk T H. Audio-visual multimedia quality assessment: a comprehensive survey. IEEE Access, 2017, 5: 21090–21117
428 Yu X, Bampis C G, Gupta P, et al. Predicting the quality of images compressed after distortion in two steps. IEEE Trans Image Process, 2019, 28: 5757–5770
429 Gu K, Zhai G, Lin W, et al. The analysis of image contrast: from quality assessment to automatic enhancement. IEEE Trans Cyber, 2015, 46: 284–297
430 Fang Y, Ma K, Wang Z, et al. No-reference quality assessment of contrast-distorted images based on natural scene statistics. IEEE Signal Process Lett, 2014, 22: 838–842
431 Gu K, Lin W, Zhai G, et al. No-reference quality metric of contrast-distorted images based on information maximization. IEEE Trans Cyber, 2016, 47: 4559–4565
432 Liu M, Gu K, Zhai G, et al. Perceptual reduced-reference visual quality assessment for contrast alteration. IEEE Trans Broadcast, 2017, 63: 71–81
433 Krasula L, Le Callet P, Fliegel K, et al. Quality assessment of sharpened images: challenges, methodology, and objective metrics. IEEE Trans Image Process, 2017, 26: 1496–1508
434 Deng Y, Loy C C, Tang X. Image aesthetic assessment: an experimental survey. IEEE Signal Process Mag, 2017, 34: 80–106
435 Wu L, Jin X, Zhao G, et al. Two open-source projects for image aesthetic quality assessment. Sci China Inf Sci, 2019, 62: 027101
436 Guo G J, Wang H Z, Yan Y, et al. Large margin deep embedding for aesthetic image classification. Sci China Inf Sci, 2020, 63: 119101
437 Autrusseau F, Stutz T, Pankajakshan V. Subjective quality assessment of selective encryption techniques. 2010.

Subjective database. http://ivc.univ-nantes.fr/en/databases/Selective_Encryption/
438 Yue G, Hou C, Gu K, et al. No-reference quality evaluator of transparently encrypted images. IEEE Trans Mult, 2019, 21: 2184–2194
439 Adhikarla V K, Vinkler M, Sumin D, et al. Towards a quality metric for dense light fields. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017. 58–67
440 Min X, Zhou J, Zhai G, et al. A metric for light field reconstruction, compression, and display quality evaluation. IEEE Trans Image Process, 2020, 29: 3790–3804
441 Viola I, Ebrahimi T. Valid: visual quality assessment for light field images dataset. In: Proceedings of IEEE International Conference on Quality of Multimedia Experience, 2018. 1–3
442 Gupta P, Sinno Z, Glover J L, et al. Predicting detection performance on security X-ray images as a function of image quality. IEEE Trans Image Process, 2019, 28: 3328–3342
443 Hu M, Zhai G, Xie R, et al. A wavelet-predominant algorithm can evaluate quality of THz security image and identify its usability. IEEE Trans Broadcast, 2019. doi: 10.1109/TBC.2019.2901388
444 Hu M, Min X, Zhu W, et al. Terahertz security image quality assessment by no-reference model observers. In: Proceedings of International Forum on Digital TV and Wireless Multimedia Communications, 2017. 100–114
445 Chen W, Gu K, Min X, et al. Partial-reference sonar image quality assessment for underwater transmission. IEEE Trans Aerosp Electron Syst, 2018, 54: 2776–2787
446 Chen W, Gu K, Lin W, et al. Reference-free quality assessment of sonar images via contour degradation measurement. IEEE Trans Image Process, 2019, 28: 5336–5351
447 Yan Z, Liu Q, Zhang T, et al. Exploring QoE for power efficiency: a field study on mobile videos with lcd displays. In: Proceedings of ACM International Conference on Multimedia, 2015. 431–440
448 Sun W, Zhai G, Min X, et al. Dynamic backlight scaling considering ambient luminance for mobile energy saving. In: Proceedings of IEEE International Conference on Multimedia and Expo, 2017. 25–30
449 Video Quality Experts Group (VQEG). Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. https://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-ii/frtv-phase-ii.aspx
450 Tanchenko A. Visual-PSNR measure of image quality. J Visual Commun Image Represent, 2014, 25: 874–878
451 Chang H, Zhang Q, Wu Q, et al. Perceptual image quality assessment by independent feature detector. Neurocomputing, 2015, 151: 1142–1152
452 Wang S, Deng C, Lin W, et al. NMF-based image quality assessment using extreme learning machine. IEEE Trans Cybern, 2017, 47: 232–243
453 Liu D, Xu Y, Quan Y, et al. Reduced reference image quality assessment using regularity of phase congruency. Signal Process Image Commun, 2014, 29: 844–855
454 Zhou W, Zhang S, Pan T, et al. Blind 3D image quality assessment based on self-similarity of binocular features. Neurocomputing, 2016, 224: 128–134