Recent Advances and Challenges in Video Quality Assessment

LI Dingquan, JIANG Tingting, and JIANG Ming

(Peking University, Beijing 100871, China)

Abstract: Video quality assessment (VQA) plays a vital role in the field of video processing, including areas of video acquisition, video filtering in retrieval, video compression, video restoration, and video enhancement. Since VQA has gained much attention in recent years, this paper gives an up-to-date review of VQA research and highlights current challenges in this filed. The subjective study and common VQA databases are first reviewed. Then, a survey on the objective VQA methods, including full-reference, reduced-reference, and no-reference VQA, is reported. Last but most importantly, the key limitations of current research and several challenges in the field of VQA are discussed, which include the impact of video content, memory effects, computational efficiency, personalized video quality prediction, and quality assessment of newly emerged videos.

Keywords: databases; perceptual optimization; personalization; video content; VQA

DOI: 10.12142/ZTECOM.201901002

http://kns.cnki.net/kcms/detail/34.1294.TN.20190305.1714.002.html, published online March 5, 2019

Manuscript received: 2018-06-09

1 Introduction

In recent years, video-based applications (e.g., video conferencing, video surveillance, and digital television) are growing rapidly in all walks of life. Especially, with the evolution of network and video technologies, people can capture videos to record their daily life with portable mobile devices wherever and whenever they like, and share the interesting ones with other people through social networking services. There is no doubt that video traffic has been the largest part of Internet traffic. However, videos pass through several processing stages before they finally reach the end users of the videos (typically the human consumers). Most of these stages impair the perceived video quality, while some of them try to improve the perceived video quality. Therefore, to provide a satisfying end-user experience, video quality assessment (VQA) is a crucial step in many video-based applications. VQA has many practical applications, including quality monitoring in real time; performance evaluation of video systems for video acquisition, compression, transmission, enhancement, display, and so on; and perceptual optimization of video systems.

VQA can be achieved by subjective VQA or objective VQA. The most reliable way to assess the perceived video quality is subjective VQA, which asks the subjects to rate the perceived video quality and processes the subjective ratings to obtain the overall video quality score. However, it is hard to carry out the subjective study in real-time video-based applications, since the subjective experiments are inconvenient, expensive, and inefficient. To automatically predict perceived video quality in real-time video-based applications, we need efficient and effective objective VQA methods.

Subjective VQA is still necessary since we need to benchmark the objective VQA methods with the "ground truth" provided by the subjective VQA, although it has so many drawbacks. Many researchers throw themselves into subjective VQA to construct benchmarking databases. In short, one constructs a video database that can reflect the variety of video

This work was partially supported by National Basic Research Program of China ("973" Program) (2015CB351803), the National Natural Science Foundation of China (61390514, 61527804, 61572042, 61520106004), Sino-German Center (GZ 1025). We also acknowledge the High-Performance Computing Platform of Peking University for providing computational resources.

content and distortions in the considered applications, and the conducted subjective study enables the constructed video database to be a benchmarking VQA database.

Developing objective VQA methods that correlate well with subjective VQA is the main goal of VQA research. According to the availability of reference videos, objective VQA methods include three types: full-reference VQA (FR-VQA), reduced-reference VQA (RR-VQA), and no-reference VQA (NR-VQA). FR-VQA methods, such as motion-based video integrity evaluation (MOVIE) index [1], require the distorted video and the corresponding pristine reference video as well. The complete access to the reference video accounts for the excellent performance of FR-VQA methods since FR-VQA can be seen as the fidelity measure. RR-VQA methods, such as spatio-temporal reduced reference entropic differences (ST-RRED) [2], lie somewhere between FR-VQA and NR-VQA, and only need partial information of the reference video in addition to the distorted one. Compared to FR-VQA, RR-VQA can achieve a good tradeoff between bandwidth occupation and superb performance. NR-VQA methods, such as video intrinsic integrity and distortion evaluation oracle (VIIDEO) [3], predict the perceived video quality without any access to the reference video. Since the reference videos are unavailable in most practical applications, NR-VQA is preferable but also more challenging.

The research field of VQA is in a rapid growth, with the fact that more and more works on new VQA methods, extensions of existing ones, and applications of these VQA methods to other disciplines are put forward every year. The goal of this paper is to provide an up-to-date review of the recent advances of VQA research as a complement to the previous reviews in [4] and [5], and more importantly to highlight the current challenges in this research field. Based on the overview of recent VQA methods, we discuss key limitations of the current VQA research and highlight some challenges in the field of VQA research that we are facing nowadays, including but not limited to the impact of video content, the memory effects and long-term dependencies, the computational efficiency and memory efficiency, the personalized video quality prediction, and the quality assessment of newly emerged videos (e.g., high dynamic range (HDR) panoramic videos) as well as quality assessment guided perceptual optimization of video systems.

This paper is organized as follows. A briefly review on the subjective VQA and public benchmarking VQA databases is presented in Section 2. Section 3 reviews the recent proposed objective VQA methods including FR-VQA, RR-VQA and NR-VQA methods. The key limitations of current VQA research and the challenges in developing effective and efficient VQA methods are discussed in Section 4. Finally, we have a concluding remark in Section 5.

2 Subjective Study and VQA Databases

Subjective video quality, i.e., the video quality perceived by humans, is the most accurate estimation of video quality since humans are the ultimate video receivers. To collect the subjective video quality scores, one must first construct a video database that can reflect the "real distribution" of videos in the application, ensuring the content diversity and distortion (level and type) diversity. Then he can select a suitable method to conduct the subjective study on the database.

The ITU [6] provides the standard settings for the subjective study of video quality. There are many subjective study methods to collect the subjective ratings, including the single-stimulus (SS) and absolute category rating (ACR) method; ACR with hidden reference (ACR-HR); double stimulus impairment scale (DSIS); double stimulus continuous quality scale (DSCQS); pair comparison (PC); subjective assessment of multimedia video quality (SAMVIQ); single stimulus continuous quality evaluation (SSCQE). PC can provide more reliable subjective quality. However, in terms of the number of videos n, its time complexity is $O(n^2)$, while the complexity of other methods is only $O(n)$. So some researchers have devoted themselves to improve the PC method by HodgeRank on random graphs [7], active sampling [8], etc.

Table 1 summarizes some common VQA databases [9]–[18] with the information about the number of reference/distorted videos, distortion types, score types, and the chosen subjective study methods. More VQA databases can be found in a collection of image and video resources on the Winkler's website [19]. The distorted videos in the first six VQA databases are all obtained by applying compression and transmission errors to the reference videos, and we refer the distortions in these videos as simulated distortions, since we can reproduce exactly the same distorted videos. However, the last four VQA databases contain no reference videos, and the distorted videos in them are authentically distorted, by which we mean that we cannot easily reproduce the same distorted videos. Actually, the simulated distortions are induced by post-processing, while the authentic ones are already induced during the video capture process. The traditional VQA databases have been analyzed in previous literatures, such as [20]. Here, we give more information about the last four VQA databases that include authentic distortions.

Camera Video Database (CVD2014) [15] includes complex authentic distortions induced during the video acquisition process. It contains 234 videos of resolution 640×480 or 1 280×720 recorded by 78 different cameras. In addition to the video quality, the conductors also ask the subjects to give ratings about sharpness, graininess, color balance, darkness, and jerkiness. One should know that, unlike previous databases, CVD2014 enables the audios in the videos. The database provides the raw subjective ratings, which means all the ratings from different subjects are available. The realigned MOS ranges from −6.50 to 93.38.

LIVE-Qualcomm Subjective Mobile In-Capture Video Quality Database [16] aims at authentic, in-capture video distortions

▼Table 1. VQA databases with the subjective study methods, numbers of (#) reference/distorted videos and score types

VQA Database	Subjective Study Method	#Reference/Distorted Videos	Score Type**
VQEG FR-TV Phase I [9]	DSCQS	22/352	DMOS+σ
VQEG HDTV [10]	ACR-HR	49/740	Raw
EPFL-PoliMI [11]	SS/ACR	12/156	Raw
LIVE [12]	ACR-HR	10/150	DMOS+σ
LIVE Mobile [13]	SSCQE-HR	10/200	DMOS+σ
CSIQ [14]	SAMVIQ	12/216	DMOS+σ
CVD2014 [15]	SS/ACR	None/234	Raw
LIVE-Qualcomm [16]	SS/ACR	None/208	MOS+σ
KoNViD-1k* [17]	SS/ACR	None/1 200	Raw
LIVE-VQC* [18]	SS/ACR	None/585	MOS+σ

*The subjective study of KoNViD-1k and LIVE-VQC is conducted on the crowdsourcing platform.
**σ indicates the standard deviation of subjective rating and "raw" means that all subjective data are available.

ACR: absolute category rating
ACR-HR: ACR with hidden reference
CSIQ: Computational and Subjective Image Quality
CVD2014: Camera Video Database
DMOS: difference mean opinion score
DSCQS: double stimulus continuous quality scale
EPFL-PoliMI: École Polytechnique Fédérale de Lausanne and Politecnico di Milano
FR-TV: full-reference television
HDTV: high definition television
KoNViD-1k: Konstanz Natural Video Database
LIVE: Laboratory for Image & Video Engineering
LIVE-VQC: LIVE Video Quality Challenge Database
MOS: mean opinion score
SAMVIQ: subjective assessment of multimedia video quality
SS: single stimulus
SSCQE: single stimulus continuous quality evaluation
VQA: video quality assessment
VQEG: Video Quality Experts Group

since the simulated distortions in previous databases cannot reflect these in-capture distortions. It consists of 208 videos of resolution 1 920×1 080 captured by eight different smart-phones and models six in-capture distortions (artifacts, color, exposure, focus, sharpness, and stabilization). The subjective study is carried out on 39 subjects, and the realigned MOS ranges from 16.5621 to 73.6428.

Konstanz Natural Video Database (KoNViD-1k) [17] focuses on authentic distortions "in the wild". To guarantee the video content diversity, it comprises a total of 1200 videos of resolution 960×540 that are fairly sampled from a large public video dataset, YFCC100M. In terms of the video content diversity, KoNViD-1k is now the largest VQA database in the community. The large scale subjective study is not suitable to be conducted in the laboratory environments, so the crowdsourcing platform is chosen. KoNViD-1k also provides the raw data of the subjective study, and the MOS ranges from 1.22 to 4.64.

LIVE Video Quality Challenge Database (LIVE-VQC) [18] is another VQA database including authentic distortions "in the wild". Same as KoNViD-1k, the large-scale study of LIVE-VQC is also conducted on the crowdsourcing platform. The subjective study has 4 776 unique participants, yielding more than 205 000 opinion scores on the 585 videos.

3 Objective Video Quality Assessment

In 2011, FR-VQA and RR-VQA methods were classified and reviewed [4], while Shahid et al. [5] classified and reviewed NR-VQA methods three years later. The research of VQA is in a rapid growth, and it has gained more attention in recent years. There have been a lot of newly proposed VQA methods since the two review articles published, thus an up-to-date review of the recent progress in VQA research is needed. Here, we give an overview of the recent advances of FR-VQA, RR-VQA, and NR-VQA methods in the following three subsections.

3.1 Full-Reference Video Quality Assessment

The research of FR-VQA methods has a long history. Since the FR-VQA methods have full access to the reference information, they can usually achieve an acceptable performance. Structural information is proved to be essential for image quality assessment (IQA), so it should be also useful for VQA. Different from images, videos have one more dimension over the time axis. So motion information should also be crucial for VQA. Furthermore, to develop an FR-VQA method that correlates well with human perception, investigating the knowledge of human visual system (HVS) is very helpful. We roughly classify the FR-VQA methods into three categories, i.e., structural information guided methods, motion information tuned methods, and HVS inspired perceptual hybrid methods.

(1) Structural information guided methods: Due to the success of structural similarity (SSIM) [21] in the field of IQA, some works in the field of VQA exploit the structural information. The most direct work that extends SSIM to video domain is proposed in [22]. Wang and Li [23] consider the frame-wise SSIM with motion associated weighting, where the motion information is obtained from a statistical model of human visual speed perception. With a novel concept of motion vector reuse, Moorthy and Bovik propose an efficient FR-VQA method, called the motion compensated SSIM (MC-SSIM) [24]. In [25], hysteresis effect is found in the subjective study, so temporal hysteresis pooling is applied to frame-wise SSIM, which is proved to be better than simply taking an average. Wang et al. extract structural information from local spatial-temporal regions [26]. More specifically, the structural information in the local space-time region is represented by the largest eigenvalue and its corresponding eigenvector of the 3D structure tensor. Besides luminance, contrast, structure similarity, Xu et al. consider the space-temporal texture by a rotation sensitive 3D texture pattern [27]. Motivated by the contrast effect, they refine the frame quality score based on the score of the previous frame. In [28], Park et al. propose a video quality pooling method to pool the frame-wise SSIM scores, which emphasizes the "worst" scores in the space-time regions.

(2) Motion information tuned methods: Motion information is very important in the videos, and this encourages developing VQA methods that utilize motion information. Seshadrinathan and Bovik put forward the MOVIE index, an FR-VQA method that considers motion perception [1]. It captures spatial distortions by spatial MOVIE maps and temporal distortions by temporal MOIVE maps, where the temporal MOVIE index is calculated with the guide of additional motion vector information. Vu et al. extend the most apparent distortion (MAD) index [29] to the video domain by taking into account of human perception on motion distortions, resulting the spatial-temporal MAD (ST-MAD) method for FR-VQA [30]. Finding that distortions can affect local optical flow statistics, Manasa and Channappayya measure the amount of distortions by the deviations of these statistics from the pristine optical flow statistics [31]. Yan and Mou [33] decompose the spatiotemporal slice images into simple motion areas and complex motion areas, and then use gradient magnitude standard deviation (GMSD) [32] to estimate the distortions in these two parts.

(3) HVS inspired perceptual hybrid methods: The goal of objective VQA is to predict video quality that correlates well with human perception, so HVS mechanism can inspire new ideas on VQA. Aydin et al. [34] propose an FR-VQA method that considers luminance adaptation, spatiotemporal contrast sensitivity and visual masking. Taking distortion detection and visual masking effects into account, Zhang and Bull [35] exploit noticeable distortion and blurring artifacts, and predict video quality by adaptively combining these two terms through a nonlinear model. Visual attention is also an important part of HVS, so some works have tried to investigate the impact of visual saliency or its implications in the field of VQA [36]–[38]. Based on the fact that HVS has the property of energy compaction representation, He et al. [39] propose an FR-VQA method by transforming the videos into the 3D discrete cosine transform (3D-DCT) domain and exploiting the energy and frequency distribution with statistical models. In [40] and [41], several perceptual-related features and methods are combined to boost the performance. Recently, in [42], video multi-method assessment fusion (VMAF) [41] is extended to embedding effective temporal features, and the resulting two methods, called spatiotemporal VMAF (ST-VMAF) and ensemble VMAF (E-VMAF), show further improvement over the VMAF method.

3.2 Reduced-Reference Video Quality Assessment

Although FR-VQA methods have the most promising performance, they have limited applications since the original videos are usually unavailable in many real-world video applications. On the other hand, NR-VQA is an extremely difficult task since it does not have access to the reference information at all. These call for a tradeoff between FR-VQA and NR-VQA tasks, and RR-VQA aims to provide this compromise.

The goal of RR-VQA methods is to reduce the issue of high bandwidth occupation in FR-VQA with minor sacrifice of performance. Video quality model (VQM) is an RR-VQA method that first calibrates the reference video and the distorted video then extract low-bandwidth spatial and temporal features to predict video quality [43]. It only requires reference data of around 4% of the size of the uncompressed video sequence, which makes it possible to perform real-time in-service quality measurements. For video quality monitoring applications, Masry et al. [44] exploit the multichannel decomposition of videos using wavelet transform with a coefficient selection mechanism that allows to adjust the bitrate of the reference video decomposition. The reference bitrates can be as low as 10 kbit/s while the proposed method keeps a good performance [44]. Gunawan and Ghanbari [45] propose an RR-VQA for compressed videos based on harmonics gain and loss information created by a discriminative analysis of harmonic strength computed from edge-detected images. Without explicit motion estimation process, Zeng and Wang [46] directly examine temporal variations of local phase structures for RR-VQA in the complex wavelet transform domain. The resulting method is very easy to be adopted by real-world video communication systems since it has only five features with very low rate of reference information. Based on the analysis of contrast and motion sensitivity characteristics of HVS, Wang et al. [47] propose a spatiotemporal information selection mechanism for RR-VQA to reduce the rate of reference information needed.

It is an issue for RR-VQA how to integrate features over time axis. To predict video quality, Le Callet et al. [48] combine three types of perceptual features (frequency content, temporal content, and blocking effects) using a time-delay neural network. It should be noted that the proposed method requires the subjective scores provided by the SSCQE method. Zhu et al. [49] propose a practical strategy for optimizing feature integration, which includes a linear model for local alignment and a non-linear model for quality calibration.

In recent years, the research of RR-VQA has been considering natural scene statistics (NSS) since distortions can alter the statistical regularities related to scene, i.e., change the NSS, evidenced by IQA methods, e.g. naturalness image quality evaluator (NIQE) [50]. Ma et al. [51] develop an RR-VQA method that exploits spatial information loss with an energy variation descriptor and exploits temporal information loss with temporal characteristics of the inter-frame histogram modeled by a statistical model. Soundararajan and Bovik [2] consider using Gaussian scale mixture model to model the wavelet coefficients of frames and frame differences, and then the measured spatial and temporal information differences between the reference and distorted videos are combined to predict video quality. ST-RRED [2] is shown to have robust performance over a wide range of VQA datasets. To further reduce complexity without sacrificing performance, Bampis et al. [52] propose the spatial efficient entropic differencing for quality assessment (SpEED-QA). Like NIQE but unlike ST-RRED, SpEED-QA applies NSS model in the spatial domain, and calculates local entropic differencing between reference and distorted videos. Since it does not need to wavelet transform, SpEED-QA is much faster than ST-RRED.

3.3 No-Reference Video Quality Assessment

In most practical video applications, the pristine videos are unavailable. For example, during the video capture process, it is incapable to capture "perfect" videos which are totally free of distortions. The additional information of the reference video also leads to high bandwidth occupation during video transmission. Moreover, people can perceive the video quality without a reference video. Therefore, NR-VQA is a more natural and preferable way to assess the perceived video quality. Over the years, numerous efforts have been put into studying distortion-specific NR-VQA methods which make assumptions on the distortion type. These methods focus on estimating the perceived quality of videos with specific distortions, such as H. 264/AVC compression [53], transmission error [54], [55], exposure distortion [56], channel-induced distortion [57], shakiness [58], spatially correlated noise [59], and scaling artifacts [60]. However, less efforts have been put into developing non-distortion-specific NR-VQA methods. This is because non-distortion-specific NR-VQA is more general and challenging since it is unaware of distortion types. With the development and applications of machine learning in the field of VQA, non-distortion-specific NR-VQA has gained much attention in recent years. Here, we give an overview of the recent advances in developing non-distortion-specific NR-VQA methods.

Some works extract frame-wise features and pool them over the time axis to obtain the video-level features for quality assessment. Xu et al. propose an NR-VQA method, called V-CORNIA, which is based on unsupervised feature learning [61]. Spatial features are first extracted in a frame-wise way based on a modification of CORNIA [62] with the max-min pooling strategy. Then, a support vector regression (SVR) model taking these features as inputs is trained for approximating frame quality to GMSD [32]. Hysteresis pooling [25] is finally employed to pool the frame quality over temporal axis. Men et al. use contrast, blurriness [63], colorfulness, spatial information (SI) and temporal information (TI) for quality assessment since they are important attributes that are related to the perceived video quality [64]. The video-level features are represented by the average of these frame-level attributes over temporal axis, and a feature combination model is proposed to map the five attributes to video quality.

Some works further consider the information contained in two adjacent frames, e.g., statistics of frame differences and optical flow. Saad et al. develop an NR-VQA method, known as VBLIINDS, which makes use of three types of features: spatial-temporal features based on a natural video statistics (NVS) model of frame differences in DCT domain, spatial naturalness index using NIQE [50], and motion-related features, i.e., motion coherency and ego-motion [65]. Finally, these features are mapped to video quality predictions by training an SVR model with linear kernel. Manasa and Channappayya propose an NR-VQA method, named FLOSIM-FR, which is based on the optical flow irregularities induced by distortions [66]. Besides, the intra-patch and inter-patch irregularities are measured in a frame-wise way while the distortion-induced flow randomness and frame irregularities are measured based on consecutive frames. The mapping between the extracted features and the video quality score is also achieved by an SVR model. Unlike the above methods, [3] and [67] are free of both distortion types and subjective ratings, which belong to "opinion-free" methods. Mittal et al. develop an efficient NR-VQA method, named VIIDEO, which is based on quantifying the intrinsic statistical irregularities due to the existence of distortions and examining the inter-subband correlations for quality assessment in the frame-difference domains [3]. By considering internal generative mechanism of HVS, Zhu et al. propose a complete blind VQA method based on spatio-temporal internal generative mechanism (ST-IGM) [67]. This method first decomposes the video content into the predicted part and the uncertain part by applying a spatio-temporal autoregressive prediction model on adjacent frames, then employs an improved NVS model to evaluate the quality of these two parts, and finally combines the two quality scores with a weighted geometric mean.

The other works directly consider cubes of video slices to exploit the spatial, temporal, and spatio-temporal information simultaneously. Li et al. develop an NR-VQA method based on an NVS model in the 3D-DCT domain, where the NVS features of short video clips (of size 4×4×4) are extracted according to the statistical analysis on basic spectral behavior, NVS shape parameter, energy fluctuation, and distribution variation [68]. These NVS features are then pooled over temporal axis to get the video-level features, and the principal components of the video-level features are fed into a linear SVR model for quality prediction. Li et al. [69] propose shearlet- and CNN-based NR VQA (SACONVA), an NR-VQA method based on 3D shearlet transform and 1D convolutional neural network (CNN). 3D shearlet transform is first employed to extract primary spatio-temporal features for video clips of size 128×128×128. 1D CNN is used for exaggerating and integrating discriminative parts of the primary features, followed by a logistic regression for video quality prediction as well as a softmax classification layer for video distortion classification. Shabeer et al. [70] extract spatio-temporal features by modelling the coefficients of sparse representation of video slices, where the spatio-temporal dictionaries are first constructed by the popular k-singular value decomposition (K-SVD) algorithm.

4 Challenges

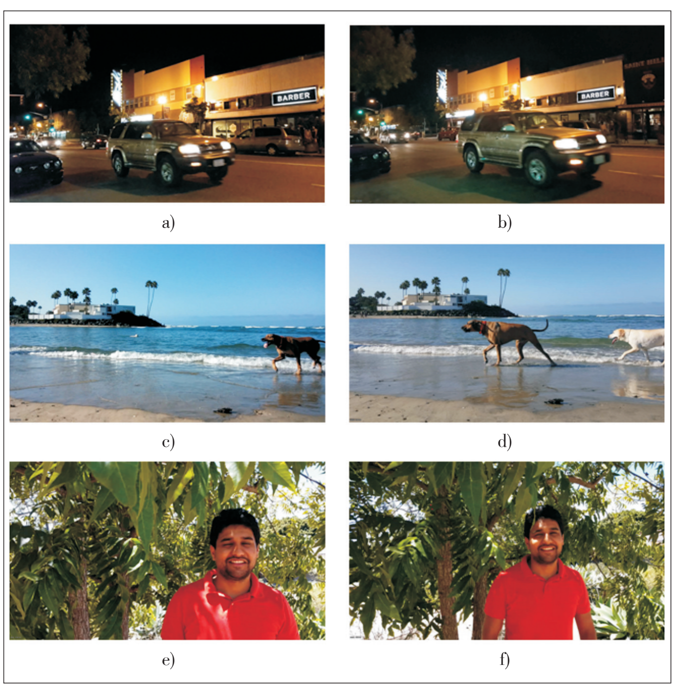
Although the previous two sections show that great progress has been made in the field of VQA research, there still remains some challenges in bridging the gap between human perception and objective VQA. In this section, we discuss several challenging issues, all of which are important aspects for overcoming barriers on the road of developing objective VQA methods that correlate well with human perception.

4.1 Impact of Video Content

The video content diversity has a strong impact on the estimation of perceived video quality since the occurrence probability of distortions and the human tolerance thresholds for distortions vary in different video content/scenes. **Fig. 1** shows an example, where the six videos suffer from almost the same levels of distortion artifacts. However, the two different videos with the same video content ("NightSence", "DogsOnBeach", or "ManUnderTree") have similar perceived quality, while the two videos with different video content have very different perceived quality. Specifically, comparing Figs. 1a/1b to Figs. 1c/1d, we can see that "DogsOnBeach" has significantly higher MOS values than "NightScene". This is because humans tend to give a higher rating for day scene videos, compared to night scene videos. Comparing Figs. 1e/1f to Figs. 1c/1d, we can see that "DogsOnBeach" has higher MOS values than "ManUnderTree". This is because humans are more sensitive to distortions occurred in human videos than in landscape videos. The depicted examples support that video content can affect human perception on the perceived quality of distorted videos.

Most of the existing objective VQA methods do not fully take the video content information into account, which may cause the performance decline when the VQA methods are tested on cross-content videos and thus cannot meet the requirements of real-world video applications that contain abundant video content information. In the FR-VQA tasks, the reference video contains the true video content information, therefore, the FR-VQA methods usually have better generalization capability on cross-content videos than the NR-VQA ones. The impact of video content on FR-VQA methods depends on how these methods utilize the reference information. Some works [36]–[38] focus on integrating visual saliency information into the VQA methods. These methods somehow further take the video content into account, since responses of human visual attention rely on "salient" video content and other salient information. NR-VQA tasks do not have the information of reference videos, thus they suffer a lot from the impact of video content. To bridge the gap between FR-VQA and NR-VQA, the first problem to be solved is finding a solution of embedding the video content information into NR-VQA.

Only NR-VQA methods are applicable to quality assessment of videos with authentic distortions. The impact of video content on quality assessment of authentically distorted videos is stronger than quality assessment of simulated distorted videos, which is evidenced by the poor performance of state-of-the-art NR-VQA methods on authentically distorted videos [15]–[18]. To bridge the gap between NR-VQA and human perception, the video content effects must be considered.



▲Figure 1. The six images are the first frame of six diferent distorted videos in LIVE-Qualcomm [16]. The distortions in these videos are at similar levels: a) NightSence, mean opinion socre (MOS)=35.7252; b) NightSence, MOS=32.2755; c) DogsOnBeach, MOS=65.9423; d) DogsOnBeach, MOS=61.2497; e) ManUnderTree, MOS=52.5666; f) ManUnderTree, MOS=57.2888. It can be seen that the perceived video quality does strongly depend on video content.

4.2 Memory Effects and Long-Term Dependencies

There exist memory effects of subjects during subjective VQA experiments, i.e., the memory of poor quality frames in the past causes subjects to provide lower quality scores for the following frames, even when the frame quality returns to acceptable levels after a time [25]. This is the evidence that long-term dependencies should be considered in the field of VQA. The existing methods consider relationships in limited adjacent frames and cannot handle the long-term dependencies well. From IQA to VQA, it is an open problem on how to deal with the memory effects and long-term dependencies in objective VQA methods.

4.3 Efficiency

Objective VQA methods can be used in real-world video applications only when they are effective and efficient. Most works focus on developing effective methods that have high performance, but less works aim at developing efficient methods which can run fast and even can be deployed in real-time video applications. Even with a C++ implementation, MOVIE [1] spends 11 438 s (more than three hours) for estimating quality of videos in LIVE [12], where the running environment is Windows 7 with 16 GB RAM and a 3.40 GHz Intel Core i7 processor [31]. This computational speed is far from the requirements in real-time applications.

Besides the computational efficiency, the memory efficiency is also a problem. RR-VQA is a way to improve memory efficiency and reduce bandwidth occupation. However, one should also pay attention to improving memory efficiency in the algorithm level if he wants his VQA method to be deployed in memory-limited applications.

4.4 Personalized Video Quality Prediction

MOS, representing the video quality given by the "average user", is not a suitable representation of video quality, since there is no "average user" in reality [71]. The standard deviation of subjective ratings may be large due to different users' preferences. Although the quality distribution of subjects can give more information about subjective quality of the video perceived by humans, it still cannot reflect the personalized video quality, which is very important for the next generation of multimedia services. The perceived video quality varies from subjects to subjects. To provide a satisfying quality of experience (QoE) for each user, personalized video quality prediction is required for guiding the user-based video delivery optimization.

The subjective studies conducted in the laboratory environments only include limited number of subjects, which is not suitable for studying the personalized video quality, and it calls for crowdsourcing platforms to collect subjective ratings from various subjects. The subjective studies should collect the user factors of each subject, including physiological factors (e.g., visual acuity and color blindness), socio-cultural factors (e.g., educational and socio-cultural background), demographics (e.g., age, sex and nationality), and psychological factors (e.g., mood and interest). Besides the environments and subjects of subjective studies, the materials, i.e., the constructed video databases, are required to contain enough video content to reflect the real distribution of videos in the applications.

The ultimate goal of personalized video quality prediction is to achieve user-centered optimization and adaptation of video applications. Quantifying the individual differences/preferences and embedding them into the VQA methods to reflect the personalized video quality are challenging but will be desired in the next generation multimedia services.

4.5 Quality Assessment of Newly Emerged Videos and Its Applications

With the development of digital devices and multimedia services, there are many emerging videos. These new videos have some new characteristics, which may raise new challenges for quality assessment research. Stereoscopic 3D video quality assessment needs to further consider depth perception [72]; VQA methods for low/standard dynamic range videos cannot directly be used for HDR videos due to different dynamic ranges [73]; quality assessment of panoramic videos/first-person videos/free-view point videos are needed with the development and popularization of virtual reality technology [74]–[76]; etc. The emerging videos become more and more popular, and they call for new VQA methods. At the meantime, the progress on quality assessment of these new videos will encourage the development of the new videos themselves. The developed quality assessment methods can also guide the perceptual optimization of video systems, e.g., the video restoration/enhancement/compression systems of both traditional and newly emerged videos. There are a good deal of challenges and opportunities in assessing the quality of newly emerged videos as well as the quality assessment guided perceptual optimization of video systems.

5 Conclusions

In this paper, we have reviewed previous works on VQA. Remarkable progress has been made in the past decade, evidenced by a number of state-of-the-art methods (especially the full-reference ones) correlating better with subjective evaluations than traditional PSNR on synthetic distorted videos. However, FR-VQA and RR-VQA methods are not applicable to authentic distorted videos since there is no way to access the reference video, thus we need NR-VQA methods in this case. Existing NR-VQA methods fail to estimate the perceived quality of authentic distorted videos, which is the evidence that the VQA research is far from mature. Then, we discuss the key limitations and five challenges of the current VQA research. We have the following statements. First, good objective VQA methods should not only consider the distortions, but also take the video content information into account. Second, memory effects and long-term dependencies are observed in the subjective studies of VQA databases, and they should be examined in developing objective VQA methods. Third, computational efficiency and memory efficiency are still big issues of quality assessment in real-time video-based applications. Fourth, by accounting for user factors, more practical VQA methods should consider predicting the personalized video quality instead of the "average video quality" of all users. At the meantime, VQA databases should provide raw data that include the user factors of subjects, and the diversity of these databases (including video content diversity and video distortion type and level diversity) should be large enough to reflect the real video distribution in the considered applications. Fifth, it is needed to develop new VQA methods for newly emerged videos (e.g., HDR panoramic videos). We also point out that how to apply the VQA methods in the perceptual optimization of video systems remains many challenges as well as great opportunities.

References

[1] SESHADRINATHAN K, BOVIK A C. Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos [J]. IEEE Transactions on Image Processing, 2010, 19(2): 335–350. DOI: 10.1109/tip.2009.2034992
[2] SOUNDARARAJAN R, BOVIK A C. Video Quality Assessment by Reduced Reference Spatio-Temporal Entropic Differencing [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(4): 684–694. DOI: 10.1109/tcsvt.2012.2214933

[3] MITTAL A, SAAD M A, BOVIK A C. A Completely Blind Video Integrity Oracle [J]. IEEE Transactions on Image Processing, 2016, 25(1): 289– 300. DOI: 10.1109/tip.2015.2502725
[4] CHIKKERUR S, SUNDARAM V, REISSLEIN M, et al. Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison [J]. IEEE Transactions on Broadcasting, 2011, 57(2): 165– 182. DOI: 10.1109/tbc.2011.2104671
[5] SHAHID M, ROSSHOLM A, LÖVSTRÖM B, et al. No-Reference Image and Video Quality Assessment: A Classification and Review of Recent Approaches [J]. EURASIP Journal on Image and Video Processing, 2014. DOI: 10.1186/1687-5281-2014-40
[6] ITU. Radiocommunication Sector & Telecommunication Standardization Sector [EB/OL]. [2018-05-27]. https://www.itu.int/en/Pages/default.aspx
[7] XU Q Q, HUANG Q M, JIANG T T, et al. HodgeRank on Random Graphs for Subjective Video Quality Assessment [J]. IEEE Transactions on Multimedia, 2012, 14(3): 844–857. DOI: 10.1109/tmm.2012.2190924
[8] FAN Z W, JIANG T T, HUANG T J. Active Sampling Exploiting Reliable Informativeness for Subjective Image Quality Assessment Based on Pairwise Comparison [J]. IEEE Transactions on Multimedia, 2017, 19(12): 2720– 2735. DOI: 10.1109/tmm.2017.2711860
[9] VQEG. VQEG FR-TV Phase I Database [EB/OL]. (2000)[2018-05-27]. http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-i/frtv-phase-i.aspx
[10] VQEG HDTV Group. VQEG HDTV Database [EB/OL]. (2009)[2018-05-27]. http://www.its.bldrdoc.gov/vqeg/projects/hdtv/hdtv.aspx
[11] DE SIMONE F, NACCARI M, TAGLIASACCHI M, et al. Subjective Assessment of H.264/AVC Video Sequences Transmitted over a Noisy Channel [C]// International Workshop on Quality of Multimedia Experience. San Diego, USA, 2009: 204–209. DOI: 10.1109/QOMEX.2009.5246952
[12] SESHADRINATHAN K, SOUNDARARAJAN R, BOVIK A C, et al. Study of Subjective and Objective Quality Assessment of Video [J]. IEEE Transactions on Image Processing, 2010, 19(6): 1427– 1441. DOI: 10.1109/tip.2010.2042111
[13] MOORTHY A K, CHOI L K, BOVIK A C, et al. Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies [J]. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(6): 652– 671. DOI: 10.1109/jstsp.2012.2212417
[14] VU P V, CHANDLER D M. ViS3: An Algorithm for Video Quality Assessment via Analysis of Spatial and Spatiotemporal Slices [J]. Journal of Electronic Imaging, 2014, 23(1): 013016. DOI: 10.1117/1.jei.23.1.013016
[15] NUUTINEN M, VIRTANEN T, VAAHTERANOKSA M, et al. CVD2014: A Database for Evaluating No-Reference Video Quality Assessment Algorithms [J]. IEEE Transactions on Image Processing, 2016, 25(7): 3073– 3086. DOI: 10.1109/tip.2016.2562513
[16] GHADIYARAM D, PAN J, BOVIK A C, et al. In-Capture Mobile Video Distortions: A Study of Subjective Behavior and Objective Algorithms [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(9): 2061–2077. DOI: 10.1109/tcsvt.2017.2707479
[17] HOSU V, HAHN F, JENADELEH M, et al. The Konstanz Natural Video Database (KoNViD-1k) [C]//Ninth International Conference on Quality of Multimedia Experience (QoMEX). Erfurt, Germany, 2017: 1– 6. DOI: 10.1109/QoMEX.2017.7965673
[18] SINNO Z, BOVIK A C. Large Scale Study of Perceptual Video Quality [EB/OL]. (2018-03-05)[2018-05-27]. https://arxiv.org/abs/1803.01761
[19] WINKLER S. Image and Video Quality Resources [EB/OL]. [2018-05-27]. http://stefan.winkler.site/resources.html
[20] WINKLER S. Analysis of Public Image and Video Databases for Quality Assessment [J]. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(6): 616–625. DOI: 10.1109/jstsp.2012.2215007
[21] WANG Z, BOVIK A C, SHEIKH H R, et al. Image Quality Assessment: From Error Visibility to Structural Similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600–612. DOI: 10.1109/tip.2003.819861
[22] WANG Z, LU L G, BOVIK A C. Video Quality Assessment Based on Structural Distortion Measurement [J]. Signal Processing: Image Communication, 2004, 19 (2): 121–132. DOI: 10.1016/s0923-5965(03)00076-6
[23] WANG Z, LI Q. Video Quality Assessment Using a Statistical Model of Human Visual Speed Perception [J]. Journal of the Optical Society of America A, 2007, 24(12): B61. DOI: 10.1364/josaa.24.000b61
[24] MOORTHY A K, BOVIK A C. Efficient Video Quality Assessment along Temporal Trajectories [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(11): 1653–1658. DOI: 10.1109/tcsvt.2010.2087470
[25] SESHADRINATHAN K, BOVIK A C. Temporal Hysteresis Model of Time Varying Subjective Video Quality [C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Prague, Czech Republic, 2011: 1153–1156. DOI: 10.1109/ICASSP.2011.5946613
[26] WANG Y, JIANG T T, MA S W, et al. Novel Spatio-Temporal Structural Information Based Video Quality Metric [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(7): 989– 998. DOI: 10.1109/tcsvt.2012.2186745
[27] XU Q Q, WU Z P, SU L, et al. Bridging the Gap between Objective Score and Subjective Preference in Video Quality Assessment [C]//IEEE International Conference on Multimedia and Expo. Suntec City, Singapore, 2010: 908–913. DOI: 10.1109/ICME.2010.5583853
[28] PARK J, SESHADRINATHAN K, LEE S, et al. Video Quality Pooling Adaptive to Perceptual Distortion Severity [J]. IEEE Transactions on Image Processing, 2013, 22(2): 610–620. DOI: 10.1109/tip.2012.2219551
[29] CHANDLER D M. Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy [J]. Journal of Electronic Imaging, 2010, 19 (1): 011006. DOI: 10.1117/1.3267105
[30] VU P V, VU C T, CHANDLER D M. A Spatiotemporal Most-Apparent-Distortion Model for Video Quality Assessment [C]//18th IEEE International Conference on Image Processing. Brussels, Belgium, 2011: 2505–2508. DOI: 10.1109/ICIP.2011.6116171
[31] MANASA K, CHANNAPPAYYA S S. An Optical Flow-Based Full Reference Video Quality Assessment Algorithm [J]. IEEE Transactions on Image Processing, 2016, 25(6): 2480–2492. DOI: 10.1109/tip.2016.2548247
[32] XUE W F, ZHANG L, MOU X Q, et al. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index [J]. IEEE Transactions on Image Processing, 2014, 23(2): 684–695. DOI: 10.1109/tip.2013.2293423
[33] YAN P, MOU X Q. Video Quality Assessment Based on Motion Structure Partition Similarity of Spatiotemporal Slice Images [J]. Journal of Electronic Imaging, 2018, 27(3): 1. DOI: 10.1117/1.jei.27.3.033019
[34] AYDIN T O, CADÍK M, MYSZKOWSKI K, et al. Video Quality Assessment for Computer Graphics Applications [J]. ACM Transactions on Graphics, 2010, 29 (6): 1. DOI: 10.1145/1882261.1866187
[35] ZHANG F, BULL D R. A Perception-Based Hybrid Model for Video Quality Assessment [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(6): 1017–1028. DOI: 10.1109/tcsvt.2015.2428551
[36] YOU J Y, EBRAHIMI T, PERKIS A. Attention Driven Foveated Video Quality Assessment [J]. IEEE Transactions on Image Processing, 2014, 23(1): 200–213. DOI: 10.1109/tip.2013.2287611
[37] PENG P, LIAO D P, LI Z N. An Efficient Temporal Distortion Measure of Videos Based on Spacetime Texture [J]. Pattern Recognition, 2017, 70: 1–11. DOI: 10.1016/j.patcog.2017.04.031
[38] ZHANG W, LIU H T. Study of Saliency in Objective Video Quality Assessment [J]. IEEE Transactions on Image Processing, 2017, 26(3): 1275– 1288. DOI: 10.1109/tip.2017.2651410
[39] HE L H, LU W, JIA C C, et al. Video Quality Assessment by Compact Representation of Energy in 3D-DCT Domain [J]. Neurocomputing, 2017, 269: 108–116. DOI: 10.1016/j.neucom.2016.08.143
[40] FREITAS P G, AKAMINE W Y L, FARIAS M C Q. Using Multiple Spatio-Temporal Features to Estimate Video Quality [J]. Signal Processing: Image Communication, 2018, 64: 1–10. DOI: 10.1016/j.image.2018.02.010
[41] LI Z, AARON A, KATSAVOUNIDIS I, MOORTHY A, MANOHARA M. Toward A Practical Perceptual Video Quality Metric [EB/OL]. (2016-06) [2018-05-27]. https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652
[42] BAMPIS C G, LI Z, BOVIK A C. SpatioTemporal Feature Integration and Model Fusion for Full Reference Video Quality Assessment [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018: 1. DOI: 10.1109/tcsvt.2018.2868262
[43] PINSON M H, WOLF S. A New Standardized Method for Objectively Measuring Video Quality [J]. IEEE Transactions on Broadcasting, 2004, 50(3): 312–322. DOI: 10.1109/tbc.2004.834028
[44] MASRY M, HEMAMI S S, SERMADEVI Y. A Scalable Wavelet-Based Video Distortion Metric and Applications [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16(2): 260– 273. DOI: 10.1109/tcsvt.2005.861946
[45] GUNAWAN I P, GHANBARI M. Reduced-Reference Video Quality Assessment Using Discriminative Local Harmonic Strength with Motion Consideration [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18 (1): 71–83. DOI: 10.1109/tcsvt.2007.913755
[46] ZENG K, WANG Z. Temporal Motion Smoothness Measurement for Reduced-Reference Video Quality Assessment [C]//IEEE International Conference on Acoustics, Speech and Signal Processing. Dallas, USA, 2010: 1010–1013. DOI:

10.1109/ICASSP.2010.5495316
[47] WANG M M, ZHANG F, AGRAFIOTIS D. A very Low Complexity Reduced Reference Video Quality Metric Based on Spatio-Temporal Information Selection [C]//IEEE International Conference on Image Processing (ICIP). Quebec City, Canada, 2015: 571–575. DOI: 10.1109/ICIP.2015.7350863
[48] LE CALLET P, VIARD-GAUDIN C, BARBA D. A Convolutional Neural Network Approach for Objective Video Quality Assessment [J]. IEEE Transactions on Neural Networks, 2006, 17(5): 1316–1327. DOI: 10.1109/tnn.2006.879766
[49] ZHU K F, BARKOWSKY M, SHEN M M, et al. Optimizing Feature Pooling and Prediction Models of VQA Algorithms [C]//IEEE International Conference on Image Processing (ICIP). Paris, France, 2014: 541–545. DOI: 10.1109/ICIP.2014.7025108
[50] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a "Completely Blind" Image Quality Analyzer [J]. IEEE Signal Processing Letters, 2013, 20(3): 209–212. DOI: 10.1109/lsp.2012.2227726
[51] MA L, LI S N, NGAN K N. Reduced-Reference Video Quality Assessment of Compressed Video Sequences [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(10): 1441–1456. DOI: 10.1109/tcsvt.2012.2202049
[52] BAMPIS C G, GUPTA P, SOUNDARARAJAN R, et al. SpEED-QA: Spatial Efficient Entropic Differencing for Image and Video Quality [J]. IEEE Signal Processing Letters, 2017, 24(9): 1333–1337. DOI: 10.1109/lsp.2017.2726542
[53] ZHU K F, LI C Q, ASARI V, et al. No-Reference Video Quality Assessment Based on Artifact Measurement and Statistical Analysis [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(4): 533–546. DOI: 10.1109/tcsvt.2014.2363737
[54] ZHANG F, LIN W S, CHEN Z B, et al. Additive Log-Logistic Model for Networked Video Quality Assessment [J]. IEEE Transactions on Image Processing, 2013, 22(4): 1536–1547. DOI: 10.1109/tip.2012.2233486
[55] ZHAO T S, LIU Q, CHEN C W. QoE in Video Transmission: A User Experience-Driven Strategy [J]. IEEE Communications Surveys & Tutorials, 2017, 19(1): 285–302. DOI: 10.1109/comst.2016.2619982
[56] ROMANIAK P, JANOWSKI L, LESZCZUK M, et al. A no Reference Metric for the Quality Assessment of Videos Affected by Exposure Distortion [C]//IEEE International Conference on Multimedia and Expo. Barcelona, Spain, 2011: 1–6. DOI: 10.1109/ICME.2011.6011903
[57] VALENZISE G, MAGNI S, TAGLIASACCHI M, et al. No-Reference Pixel Video Quality Monitoring of Channel-Induced Distortion [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(4): 605–618. DOI: 10.1109/tcsvt.2011.2171211
[58] CUI Z X, JIANG T T. No-Reference Video Shakiness Quality Assessment [M]// CUI Z X, JIANG T T. eds. Computer Vision–ACCV 2016. Cham: Springer International Publishing, 2017: 396–411. DOI: 10.1007/978-3-319-54193-8_25
[59] CHEN C, IZADI M, KOKARAM A. A Perceptual Quality Metric for Videos Distorted by Spatially Correlated Noise [C]//ACM on Multimedia Conference. Amsterdam, The Netherlands, 2016: 1277–1285. DOI: 10.1145/2964284.2964302
[60] GHADIYARAM D, CHEN C, INGUVA S, et al. A No-Reference Video Quality Predictor for Compression and Scaling Artifacts [C]//IEEE International Conference on Image Processing (ICIP). Beijing, China, 2017: 3445–3449. DOI: 10.1109/ICIP.2017.8296922
[61] XU J T, YE P, LIU Y, et al. No-Reference Video Quality Assessment via Feature Learning [C]//IEEE International Conference on Image Processing (ICIP). Paris, France, 2014: 491–495. DOI: 10.1109/ICIP.2014.7025098
[62] YE P, KUMAR J, KANG L, et al. Unsupervised Feature Learning Framework for No-Reference Image Quality Assessment [C]//IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 1098–1105. DOI: 10.1109/CVPR.2012.6247789
[63] MEN H, LIN H H, SAUPE D. Empirical Evaluation of No-Reference VQA Methods on a Natural Video Quality Database [C]//Ninth International Conference on Quality of Multimedia Experience (QoMEX). Erfurt, Germany, 2017: 1–3. DOI: 10.1109/QoMEX.2017.7965644
[64] NARVEKAR N D, KARAM L J. A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD) [J]. IEEE Transactions on Image Processing, 2011, 20(9): 2678–2683. DOI: 10.1109/tip.2011.2131660
[65] SAAD M A, BOVIK A C, CHARRIER C. Blind Prediction of Natural Video Quality [J]. IEEE Transactions on Image Processing, 2014, 23(3): 1352–1365. DOI: 10.1109/tip.2014.2299154
[66] MANASA K, CHANNAPPAYYA S S. An Optical Flow-Based No-Reference Video Quality Assessment Algorithm [C]//IEEE International Conference on Image Processing (ICIP). Phoenix, USA, 2016: 2400–2404. DOI: 10.1109/ICIP.2016.7532789
[67] ZHU Y, WANG Y F, SHUAI Y. Blind Video Quality Assessment Based on Spatio-Temporal Internal Generative Mechanism [C]//IEEE International Conference on Image Processing (ICIP). Beijing, China, 2017: 305–309. DOI: 10.1109/ICIP.2017.8296292
[68] LI X L, GUO Q, LU X Q. Spatiotemporal Statistics for Video Quality Assessment [J]. IEEE Transactions on Image Processing, 2016, 25(7): 3329–3342. DOI: 10.1109/tip.2016.2568752
[69] LI Y M, PO L M, CHEUNG C H, et al. No-Reference Video Quality Assessment with 3D Shearlet Transform and Convolutional Neural Networks [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(6): 1044–1057. DOI: 10.1109/tcsvt.2015.2430711
[70] SHABEER P M, BHATI S, CHANNAPPAYYA S S. Modeling Sparse Spatio-Temporal Representations for No-Reference Video Quality Assessment [C]// IEEE Global Conference on Signal and Information Processing (GlobalSIP). Montreal, Canada, 2017: 1220–1224. DOI: 10.1109/GlobalSIP.2017.8309155
[71] ZHU Y, GUNTUKU S C, LIN W, GHINEA G, REDI J A, Measuring Individual Video QoE: A Survey, and Proposal for Future Directions Using Social Media [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14(2s): 30. DOI: 10.1145/3183512
[72] CHEN Z B, ZHOU W, LI W P. Blind Stereoscopic Video Quality Assessment: From Depth Perception to Overall Experience [J]. IEEE Transactions on Image Processing, 2018, 27(2): 721–734. DOI: 10.1109/tip.2017.2766780
[73] NARWARIA M, PERREIRA DA SILVA M, LE CALLET P. HDR-VQM: An Objective Quality Measure for High Dynamic Range Video [J]. Signal Processing: Image Communication, 2015, 35: 46–60. DOI: 10.1016/j.image.2015.04.009
[74] ZHANG Y X, WANG Y B, LIU F Y, et al. Subjective Panoramic Video Quality Assessment Database for Coding Applications [J]. IEEE Transactions on Broadcasting, 2018, 64(2): 461–473. DOI: 10.1109/tbc.2018.2811627
[75] BAI C, REIBMAN A R. Image Quality Assessment in First-Person Videos [J]. Journal of Visual Communication and Image Representation, 2018, 54: 123–132. DOI: 10.1016/j.jvcir.2018.05.005
[76] LING S Y, LE CALLET P. Image Quality Assessment for Free Viewpoint Video Based on Mid-Level Contours Feature [C]//IEEE International Conference on Multimedia and Expo (ICME). Hong Kong, China, 2017: 79–84. DOI: 10.1109/ICME.2017.8019431

Biographies

LI Dingquan received the double B.S. degrees in electronic science and technology & applied mathematics from Nankai University, China in 2015, and he is currently working toward the Ph.D. degree in applied mathematics at Peking University, China. He is a member of National Engineering Lab for Video Technology. His research interests include image/video quality assessment, perceptual optimization, and machine learning. He has published papers in *IEEE Transactions on Multimedia* and ACM Multimedia Conference.

JIANG Tingting (ttjiang@pku.edu.cn) received the B.S. degree in computer science from University of Science and Technology of China in 2001 and the Ph.D. degree in computer science from Duke University, USA in 2007. She is currently an associate professor of computer science at Peking University, China. Her research interests include computer vision and image/video quality assessment. She has published more than 40 papers in journals and conferences.

JIANG Ming received the B.Sc. and Ph.D. degrees in mathematics from Peking University, China in 1984 and 1989, respectively. He is a professor with Department of Information Science, School of Mathematical Science, Peking University since 2002. His research interests are mathematical and technical innovations in biomedical imaging and image processing.