


# Metrics and methods of video quality assessment: a brief review

Qiang Fan<sup>1</sup> · Wang Luo<sup>1</sup>  · Yuan Xia<sup>1</sup> · Guozhi Li<sup>1</sup> · Daojing He<sup>2</sup>

Received: 21 January 2017 / Revised: 10 April 2017 / Accepted: 17 May 2017 /  
Published online: 21 July 2017  
© Springer Science+Business Media, LLC 2017

**Abstract** With rapid development of video acquisition devices and wide applications of video data, more and more requirements are established to use video data query. video quality assessment and improvement become popular and important research issues which attract lots of researchers. The video quality can influence technique application, user experience, and application results. This paper firstly reviews research work on video query based applications. Then, various metrics of video quality assessment are reviewed according to the requirement of reference video, including full-reference metrics, no-reference metrics and reduced-reference metrics. In addition, methods of video quality assessment are reviewed by the features, which include visual features, bitstream-based or packet-based features, data features. Finally, a number of video quality improvement methods are briefly introduced.

**Keywords** Quality assessment · Assessment metrics · Assessment method · Video query

## 1 Introduction

Video data has been widely utilized in numerous applications. A number of applications, such as video summarization [84], object tracking, traffic status estimation [46], multi-view objects recognition [68], human activity recognition [77], categorization [78, 82], saliency detection [83], image segmentation [79], photo retargeting [66, 80] etc., which highly rely on video data for information retrieval [59] or advanced query, have been boosted by the deployment of various video capturing devices. In addition, the advancement of network

---

✉ Wang Luo  
luowang@163.com

<sup>1</sup> State Grid Electric Power Research Institute, NARI Group Corporation, Nanjing, China

<sup>2</sup> School of Computer Science and Software Engineering, East China Normal University, Shanghai, China

technology allows higher network bandwidth, which enables access to videos via Internet from anywhere.

High quality of collected videos is significant for reliable and stable video analysis services, also for the experience of end users. Limited by transmission resources, video data should be compressed to reduce its size before transmission. A lossy video compression algorithm may introduce artificial content, while packet losses or random noises in the transmission process [60] can damage the video quality. The distorted video content can pose significant negative effects on video processing and analyses, as well as unpleasant feelings to the end users. Consequently, the assessment [63] of video quality is crucial for evaluating captured videos.

The problem of video quality assessment (VQA) has attracted a lot of research interests. Numerous metrics and assessment methods were proposed to effectively evaluate the video quality [65] by utilizing various features, including visual characteristics, perceptual features and psychological features. Existing VQA metrics can be classified in different ways. One classification method is based on the requirement of an intact reference video, which includes Full-Reference (FR) metrics, No-Reference (NR) metrics and Reduced-Reference (RR) metrics [52]. As the FR metrics are heavily dependent on the availability of reference videos, RR and NR metrics are more appealing in practice.

VQA methods can also be classified using different perspectives. According to the principles in [9], VQA methods can either be visual characteristic based or perceptual characteristic based. Furthermore, visual-based methods can be further categorized into methods using natural statistical characteristics or visual features. Perceptual methods can be divided into frequency-domain or pixel-domain methods. In addition, the quality assessment methods can also be categorized as either subjective or objective [23]. The aim of subjective VQA is to evaluate the average vote score from the perspective of users. It is the most reliable way but the result of a subjective VQA needs to be annotated by a trained expert. Objective VQA methods allow to predict video quality in an automatic way with selected metrics. Besides the aforementioned classification methods, VQA methods can also be classified as two-dimensional (2D) oriented or three-dimensional (3D) oriented.

This paper focuses on the challenges of VQA, and latest works of VQA metrics, VQA methods, and video quality improvement. The remaining of this paper is structured as follows. In Section 2, we review applications of video query in different areas. Then, in Section 3 VQA metrics are briefly introduced. VQA methods are elucidated in Section 4 according to various features. Section 5 provides a brief review on improvement methods of video quality. A conclusion is drawn in Section 6.

## 2 Applications of video query

There is growing interest in query optimization in multimedia [62] databases both in academic and commercial field. It is significant and has much challenge to execute data analysis applications efficiently as data set sizes and formats grow continually [3]. Different multimedia object will use different types of techniques. It involves query segmentation which takes a query and simplified restructured query expressed in relational algorithm after modification due to views, security enforcements and semantic integrity control [44]. There are three optimization techniques which are widely used in multimedia databases, which include Content Based Query, Semantic Based Query and Metadata. In the following, we discuss these three related optimization techniques firstly and then we summarize the advantages and disadvantages of each technique.

## 2.1 Content Based Query (CBQ)

This is the most widely used type of optimization technique. It is the main motivation in recent research on multimedia databases [43]. Chan et al. access semantic information contained in video data by CBQ formulates queries [8]. Kwok et al. extract parts of the image and transform them into low level features on that particular data objects by using content based image query (CBIQ) [28]. Zhou et al. focus on the low level features such as color, motion, and texture index of the video based on CBQ [85]. Sadat et al. create a prototype for multimedia database with a query interface focusing on the dependencies of context and content information for effective query [44].

Similarity search is one of the important implementation in the CBQ type, which focuses on the inner part of multimedia data. It is an operation that finds changing pattern of sequences or subsequences that are similar to the given query sequence. There are two different elements in content based video similarity search [85]. One is video clip query, which focuses on getting similar clips from a large collection of videos that have been segmented into shots of similar length size at content boundaries. The other is video subsequence identification [64], which focuses on searching for existent of any part of a long database video that shares similar content to a query. Similarity queries can be classified into two categories which are whole matching (sequences to be compared have the same length) and sequence matching (involves the query sequence that is smaller and the comparison is done in the large sequence that best match the query sequence) [1]. Kriegel et al. focus on the effectiveness of similarity search in multimedia databases using multiple representations for video and try to integrate multiple video representations featured into query processing [27].

The advantages of CBQ are summarized as: it provides ease of query formulation and interpretation; and this technique is suitable and flexible for formulating queries. The disadvantages: CBQ provides data dependency between multimedia data; and users are dependency; and the performance is lacking when queries are executed on large data sets.

## 2.2 Semantic Based Query (SBQ)

Semantic query uses knowledge about the domain of relations, nature of data, and constraints related to database elements [48]. Semantic based search is defined as a technique that compares the original multimedia data to a prototypical category such as ‘vehicle’, clothes and others [73]. The semantic of the video is modeled by elements which extracted from different modalities of a video, such as visual information, auditory information, and text in the video frames [10]. There are four main issues involved in semantic query [48]. The first issue focuses on the query and scheme which should be dynamically used to select the relevant semantics for optimization without any addition search of semantic rule base. Second, a suitable mechanism should be available to combine selected semantic with the query. Third, cost analyzer is needed to evaluate the cost of equivalent queries and rank them accordingly. Lastly, a heuristic guide is needed to show the whole process in a meaningful way so that the process can be easily understood by the users.

In order to handle rich, temporal and spatial requirements of multimedia data, a visualized semantic model is used to increase the information content so that the users cognitive load is reduced [8]. Query and data centric method can be incorporated to optimize the acquisition process. Semantic query optimization is based on the semantic equivalence rather than the syntactic equivalence between different queries [47].

The advantages of SBQ are: large amount of computation time can be saved compared to the index structures because of the simpler and overlap-free characteristics; and the

technique can be easily extended to all transforms available. However, there are several disadvantages of semantic content such as it is difficult to extract automatically [17]; it cannot support queries on generalized concepts; and the retrieval is not precise enough for the process

## 2.3 Metadata

Metadata is the data or semantic information to classify the content, quality, condition and other characteristics of the data [57]. Metadata technique is a type of searching technique that the structured information describing characteristics assist users to identify digital content itself [37]. It changes the way of content search from using normal string to a conceptual level where users search for semantic contents of the data [49]. There are three famous metadata schemes which are MPEG-7, Dublin Core, and IEEE LOM.

MPEG-7 is different from other standard metadata that it can provide two types of schema which is divided into low level descriptions and high level ones [42]. Low level description is defined as color, texture, and the shape of the multimedia data while the high level one is the structural and semantic descriptions. MPEG-7 major focuses are video and images and the technology enables the CBIR. MPEG-7 enables the metadata to be used in different platform and applications [36].

The second scheme is called Dublin Core schemes which conceived author generated descriptions of Web sources [56]. Its aim is to define a standard that will encourage quality resource description and encourage interoperability between tools for resource discovery [21].

The third scheme is called IEEE standards that conform, integrate and refer open standards with the existing work in related areas [49]. Learning Object Metadata (LOM) standards specify a conceptual data scheme that defines the structure of a metadata instance [47]. IEEE LOM is divided into nine elements descriptors of learning resources, each of which is relatively independent and characterizes the resource from a separate aspect [17]. In addition, its specification and vocabularies were determined through discussion from the standpoints of both users and resource developers. The IEEE LOM uses hierarchical types of metadata description which is useful, and easy to be implemented in many levels of elements [70].

The advantages of Metadata are as following: rich metadata are very effective in assisting users to navigate and find desired content items; and high quality metadata is important for reliable and effective Web applications. While the disadvantages are: the more data become available, the harder and difficult to identify and extract metadata; the absence of meaningful metadata due to lack of users' attention in providing the information.

In summary, among these three techniques, CBQ is more suitable for video data, where the content of the video can be segmented firstly and compared using similarity search. This technique normally involves more than one technique to extract the content of multimedia data. However, this technique provides dependency between multimedia data. Also, the process involves the extraction of multimedia segments in order to retrieve the query output. SBQ focus on image data more. It involves algorithm to analyze the structure of the image covering the shape, color, and texture of the data by using low and high feature extraction. Since this technique involved the extraction of the inner part of the data, queries cannot support generalized concepts and the output retrieval is not precise for the process. Metadata is suitable for both of the data. However, users have to interact in defining the metadata. In order to have an effective application, high quality metadata is very important. Whereas the more data are available, the harder it is to identify and find the suitable metadata (Table 1).

**Table 1** Comparisons among three optimization techniques

Techniques	Applicability	Features
CBQ	Video data	More than one technique for extraction
SBQ	Image data	Cannot support generalized concepts and not precise
Metadata	Both of the data	Difficult to identify and find the suitable metadata

### 3 Metrics for video quality assessment

For FR metrics, they are evaluated by comparing the test video to its corresponding reference video. The mean squared error (MSE), peak signal-to-noise ratio (PSNR), and extended metrics based on MSE or PSNR, are commonly seen metrics used in FR VQA methods. These metrics are simple, but their performance in subjective correlation are poor. By integrating the assessment of the spatial and temporal distortion, Motion-based Video Integrity Evaluation (MOVIE), spatial MOVIE (SMOVIE), temporal MOVIE (TMOVIE) were also proposed as FR metrics [45]. The main drawback of FR metrics lies in their heavy dependence on the reference video, and the reference video must be aligned with the test video. Features of saliency [14], MSE and video content [6], distortion [20], etc. can be used for FR metrics design.

NR metrics generally works on the estimation of blocks, distortion, blur, noises, quantization errors, watermarks, bitstream, etc. They can be used to analyze the test video without any reference video. The WMBER metric, which detects the macro-block errors weighted by a saliency map and considers the characteristics of the human visual system (HVS), was proposed in [7]. Bitstream features can be analyzed to get metrics for video quality assessment. In [25], a prediction model of visual quality was developed by utilizing extracted bitstream-based features from partial least squares regression. A NR color-based metric for video quality is presented in [35]. It employs a flow tensor and a perceptual mask, which integrates spatio-temporal contrast sensitivity function and luminance sensitivity, to define the metric. In addition, other NR metrics can be achieved by estimating such features as blur [38], block artefacts [53].

RR VQA metrics can assess the video quality with partial reference information available. Discriminative local harmonic strength and motion measurement were used in the proposed metric [15], in which the harmonic gain/loss was evaluated by a harmonic analysis of the source video frames. Structural similarity-based metric, SSIM, utilizes coding tools based on the distributed source coding theory in [50]. To evaluate the quality of 3D videos query, compressed depth maps and color features are incorporated into the RR metric [18].

The comparison among three metrics is shown as follows (Table 2).

**Table 2** Comparisons among three metrics

Metrics	Differences	Existing models
FR	Depend on the reference video	MOVIE, SMOVIE, TMOVIE
NR	Without reference video	WMBER
RR	With partial reference information	SSIM

In order to show the researchers to follow the general state of art in this research area quickly, we introduces some available data sources and detailed applications in this area, which is shown as follows (Table 3).

## 4 Methods of video quality assessment

In this paper, we categorize methods of video quality assessment according to the type of features utilized for the evaluation of video quality. The features include visual features, data-based features and network transmission-related features. Visual features cover most characteristics utilized in quality assessment, and they are related to motion, content, distortion, spatial information, temporal information, visual attention, etc. Transmission-related features are relevant to the packets and bitstreams in the transmission of video data through the network.

### 4.1 Visual feature-based methods

#### (1) Motion feature-based methods

In [19], motion information was employed to filter unnecessary information in the spatial frequency domain, and a spatio-velocity contrast sensitivity function (SV-CSF) was introduced for objective video quality assessment. SV-CSF describes the relationship among contrast sensitivities, spatial frequencies and velocities of perceived stimuli. However, the SV-CSF cannot work in the spatial frequency domain directly in the filtering process. Video frames separated in spatial frequency domain should be obtained and these frames are weighted by contrast sensitivities in the SV-CSF model. Another motion estimation based method is suggested in [22]. In this work, the video quality is assessed by estimating motion quality and motion trajectory. And a Motion-based Video Integrity Evaluation (MOVIE) index is introduced based on motion estimation. The evaluation result demonstrates that the quality assessment score derived by MOVIE index is close to human subjective judgment.

Moreover, the salient motion is included in the assessment scheme suggested in [11]. In this work, features were proposed to describe the salient motion intensity and the coding artifacts intensity in the salient motion region. The experimental results exhibited that the salient-motion features can enhance the video quality assessment when blocking artifacts and blurring in the salient region, as well as temporal changes of regional intensities. The temporal motion smoothness of video sequences is measured in the proposed assessment method in [76] by temporal variations of local phase structures in the complex wavelet transform domain. The proposed method is adaptive to a wide range of video issues, including distortions, noise contamination, blurring, jittering and frame dropping, and it has a low reduced-reference data rate with lower computational cost.

#### (2) Content-based methods

In addition to encoder settings and network quality of service, the type of video content is another factor that can affect the video quality. A content-based method of video quality assessment was proposed in [24], in which a novel metric, Simplified Perceptual Quality Region (SPQR), was used a sign of video quality degradation. SPQR determines the face location of the speakers in the video, and the discrepancies of face location in the corresponding frames. And the evaluation result shows that the proposed method is a light-weight

**Table 3** Data sources

Data sources	Description	Link
Netflix Video Quality of Experience Database	Contain videos distorted by mixtures of dynamic network impairments and compression	<a href="http://live.ece.utexas.edu/research/LIVE_NFLXStudy/nflx_index.html">http://live.ece.utexas.edu/research/LIVE_NFLXStudy/nflx_index.html</a>
Qualcomm Subjective Mobile In-Capture Video Quality Database	Contain 208 ‘authentically distorted’ incapture distortions and their subjective scores	<a href="http://live.ece.utexas.edu/research/incaptureDatabase/index.html">http://live.ece.utexas.edu/research/incaptureDatabase/index.html</a>
HDR Subjective Image Quality Database	Containing HDR images processed by various algorithms and their subjective scores.	<a href="http://live.ece.utexas.edu/research/HDRDB/hdr_index.html">http://live.ece.utexas.edu/research/HDRDB/hdr_index.html</a>
Mobile Video Stall Database	Contain videos distorted with stalling events and their subjective scores.	<a href="http://live.ece.utexas.edu/research/LIVestallStudy/index.html">http://live.ece.utexas.edu/research/LIVestallStudy/index.html</a>
Wild Image Quality Challenge Database	Contain ‘authentically distorted’ images and their subjective scores.	<a href="http://live.ece.utexas.edu/research/ChallengeDB/index.html">http://live.ece.utexas.edu/research/ChallengeDB/index.html</a>
Flicker Video Database	Contain “flicker distorted” videos and their flicker visibility evaluations.	<a href="http://live.ece.utexas.edu/research/quality/live_flicker_video.html">http://live.ece.utexas.edu/research/quality/live_flicker_video.html</a>
Distorted Face Database	Contain ‘singly distorted’ frontal faces of varied severities.	<a href="http://live.ece.utexas.edu/research/quality/dfd.html">http://live.ece.utexas.edu/research/quality/dfd.html</a>
QoE Database for HTTP-based Video Streaming	Containing “upto the moment” subjective quality along with overall quality for 15 videos.	<a href="http://live.ece.utexas.edu/research/quality/TVSQ_VQA_database.html">http://live.ece.utexas.edu/research/quality/TVSQ_VQA_database.html</a>
Multiply Distorted Image Quality Database	Containing multiply distorted images and their subjective evaluations.	<a href="http://live.ece.utexas.edu/research/quality/live_multidistortedimage.html">http://live.ece.utexas.edu/research/quality/live_multidistortedimage.html</a>
3D Image Quality Database	Contain distorted images and their subjective evaluations.	<a href="http://live.ece.utexas.edu/research/quality/live_3dimage.html">http://live.ece.utexas.edu/research/quality/live_3dimage.html</a>
Mobile Video Quality Database	Subjective, behavioral, and objective studies over wireless network.	<a href="http://live.ece.utexas.edu/research/quality/live_mobile_video.html">http://live.ece.utexas.edu/research/quality/live_mobile_video.html</a>
Video Quality Database	Contain distorted videos and their subjective evaluations.	<a href="http://live.ece.utexas.edu/research/quality/live_video.html">http://live.ece.utexas.edu/research/quality/live_video.html</a>
Image Quality Database	Contain distorted images and their subjective evaluations.	<a href="http://live.ece.utexas.edu/research/quality/subjective.htm">http://live.ece.utexas.edu/research/quality/subjective.htm</a>

implementation. To assess the quality of high definition video streaming with packet losses, a quality of experience model presented in [30] utilizes the SSIM metric, temporal pooling method and content-based features to evaluate the video quality and performs well.

In the work presented in [4], the video content type is taken into consideration in the design of a reference-free video quality prediction model, in which the motion vector is utilized to extract the temporal information, while the spatial information is obtained with quantization parameters and the number of bits of frames. Then the derived metric is used to represent the content type of video sequences. And finally, the quality prediction model is built by combining the content type metric with encoding quantization parameters and network packet loss rate. The experimental result achieves an accuracy of 92.

Considering that contents in different regions play different roles in the perceptual quality of an image, a three-component image model was proposed in [32] to evaluate the video quality. In this work, gradient properties were used to classify the local image regions, then apply variable weights to structural similarity image index scores according to region. A frame-based video quality assessment algorithm [33] is thereby derived. Experimental results on the Video Quality Experts Group (VQEG) FR-TV Phase 1 test dataset show that the proposed algorithm outperforms existing video quality assessment methods.

An embedded water mark was utilized to estimate the video quality in the proposed method in [40]. The pseudo-random binary watermark is fused with the original video frames, and the similarity between the original and the extracted water mark is evaluated to assess the query quality of a video segment. The water mark can be embedded in small or large wavelet scales [67] for sensing small or substantial distortions of the video frame.

### (3) Methods based on distortion estimation

Distortion features of video can be used to estimate the video quality. To estimate the perceived visual distortion in an image or video frames, structural distortion is employed in [55]. The proposed method is performed on the video quality experts group Phase I FR-TV test data set, and the testing result demonstrates that the proposed method is computationally efficient.

Statistical distortion features were used in a no-reference method of video quality assessment presented in [72]. In this work, each frame is converted into the wavelet domain and the oriented band-pass response is generated by their decomposition [69]. Statistical distortion features are extracted with resulted sub-band coefficients and are used to construct a feature vector as a descriptor of the overall distortion of the frame. The video quality in the wavelet domain is achieved by the classification of the feature vectors across images and a score mapping. The temporal quality is evaluated with a motion-compensated method utilizing block and motion vector. And the overall quality is achieved by a pooling strategy.

### (4) Methods using spatial and temporal features

Spatial and temporal factors are usually combined to improve the assessment results of video quality. Spatial and temporal artefacts in videos are discussed in [41], and the result shows that those artefacts are correlated. Moreover, the contribution of spatial quality to the overall video quality is more than that of temporal quality. Based on aforementioned ideas, an objective quality evaluation model was proposed by combing spatial quality with temporal quality. In the computational model proposed in [39], the spatial and temporal factors were also combined by exploiting the worst-case pooling method and the variation of spatial quality along the temporal axis. The interaction between these two factors is determined by a machine learning algorithm. With the popularity of video-sharing services, web videos are variable in various aspects including content, capturing devices, resolution, etc. In order to



evaluate the quality of web videos, spatiotemporal factors integrating the features of video editing style are utilized to predict the quality of web videos in [58]. Then the task of quality evaluation can be seen as a problem of two-class classification.

#### (5) Methods utilizing visual attention

The query quality assessment model introduced in [75] was built up the attention theory, in which video quality is perceived in both local and global assessment. In this model, the attention map, which is derived by fusing several visual features that can have impacts on visual attention, is designed for optimizing local quality model to evaluate the degradations on the attended stimuli. A global quality model is formed by fusing four designed quality features. Then, a content adaptive linear fusion method is used for fusing these local and global features to assess the video quality.

#### (6) Methods exploiting block artefacts

Artefacts resulted by block-based codecs related to H.264/AVC were utilized in [31] to reach a no-reference metric for video quality assessment. Compared with a full-reference method, the suggested metric, Structural Similarity Index Metric (SSIM), produces a better result, and functions well in the real-time scenario. In [71], the cepstrum analysis is utilized to improve the estimation of blocking artefacts. A no-reference blocking artifacts metric was proposed in [53], in which the weighted evaluation of artefacts was performed in flat and edge regions. The result shows that it is highly consistent with visual perception.

## 4.2 Bitstream-based or packets-based methods

Packet loss is another issue when the video stream transfer through the limited bandwidth network channels. In the video delivery service over an IP network, the features of packet losses or bitstream of videos should be taken into consideration when evaluating the video quality. As for the quality assessment approach proposed in [74], spatial and temporal pooling were conducted with packet losses for evaluating the video quality, and the experimental results show that the packet loss-based method is more sensitive to the most annoying spatial regions and temporal segments. A saliency based query quality assessment method is proposed in [13] to evaluate the videos with packet losses, and the visual saliency of the pixel is used as a weight of the error of the corresponding pixel.

Instead of reconstructing video information, coding parameters were utilized to evaluate video query qualities in [29]. In this method, coding parameters, which include boundary strengths, quantization parameters and average bitrates, are extracted from the H.264/AVC bitstream. The accuracy of the proposed method is higher than previous methods, and its computational complexity is lower, which makes it a competitive candidate in real-time scenarios.

## 4.3 Methods utilizing data features

In video quality assessment, peak signal-to-noise (PSNR) and mean squared error (MSE) are two commonly used criteria. However, they are unable to fulfil the requirement of human subjective assessment. In order to consider the effects of visual sensitivity on video quality, structural similarity and visual statistical information sensitivity were taken into account in designing the assessment method. In [34], the proposed method introduces the phase spectrum of quaternion Fourier transforms, and the weighted saliency features are exploited in query quality assessment.

#### 4.4 3D-oriented assessment methods

With the popularity of 3D videos, query quality assessment methods for 3D videos attract a lot of researchers due to that assessment methods for 2D videos cannot be directly exploited to 3D videos. A no-reference stereoscopic video quality perception model for 3D video was suggested in [16]. It is built on four extracted factors, which are temporal variance, disparity variation in intra-frames, disparity variation in inter-frames and disparity distribution of frame boundary areas. The model is free from the depth map and the parameters can be estimated by linear regression.

3D Singular Value Decomposition (3D-SVD), which is a singular value decomposition in 3D space, is utilized by the models presented in [51] and [81] for query quality assessment of 3D videos. The model in [51] utilized the original video, and the distorted video is projected onto the singular vectors of the original video, then the quality of the video query can be evaluated by calculating the weighted differences between the reflection coefficients. For the method in [81], the image is separated into different planes based on depth values, then a global error is computed as the distance between the distorted image and the original image. In addition, depth information and motion cue are explored in 3D VQA method presented in [26]. By combining depth and motion cues, a weighting map based on PSNR and SSIM is generated for 3D VQA.

### 5 Video quality improvement

After assessing the video quality by either a subjective or an objective approach, various video quality enhancement methods can be utilized to improve the quality of experience perceived by the end users. Firstly, video quantization can be analyzed to achieve an optimal value of the quantizer\_scale factor for improving the video quality, and this method can save bandwidth in an IPTV network. It is used as an automatic measure to improve the video quality received by the end user [5]. When transmitting video frames over the network with bursting losses, the video quality can be improved by classifying video frames at the edge node and differentiating the service on the network. Optical Burst Switched (OBS) networks and Optical Packet Switched (OPS) networks are examples of such technologies and Video quality is evaluated by means of a no reference video quality metric (the Frame Starvation Ratio) [12].

Moreover, the advanced video coding (AVC) which is also called as H.264/ MPEG-4 is the latest compression standard available for video compression. Fidelity Range Extension (FRExt) is the recent work done on the AVC which includes a number of enhanced capabilities relative to the base standards. Human vision models can be incorporated into the standard codes for improving video quality, and a contrast sensitivity function is introduced in the transform coding stage of standard codes. AVC with FRExt features has been implemented first and the quality of the reconstructed video signals are evaluated using both subjective and objective measures such as MOS, PSNR, MSSIM, VIF and VSNR [2]. In the work presented in [54], a Multi Layer Streaming-Simplified DCT Domain Transcoder (MLS-SDDT) is proposed to video quality improvement, including an FGS compatible Simplified DCT Domain Transcoder (FGS-SDDT) architecture for MPEG-1/2/4 single layer transcoding and an R-D optimized multi-layer streaming model for video quality improvement. By applying the MLS-SDDT to FGS-toMPEG-1/2/4 single layer transcoding, experiments show 1.4 7.0 dB PSNR improvement for MPEG-1 and 1.9 8.6 dB improvement for MPEG-2 compared to SDDT architecture.

## 6 Conclusion

Considering the popularity of video query in various applications [61] and great importance of VQA, we conducted a comprehensive review on VQA metrics and VQA methods, as well as video quality improvement in this work. We introduced three optimization techniques which are widely used in multimedia databases, which include Content Based Query, Semantic Based Query and Metadata. Based on visual or physical features such as color, depth information, structural similarity, bitstream, distortion, video content, spatial and temporal information, etc., we gave a detailed description of FR metrics, NR metrics and RR metrics. Later we gave the classification and introduction of VQA methods and some experimental results, such as Visual feature-based methods, Bitstream-based or Packets-based Methods, etc. And finally, a brief introduction of video quality improvement methods is given.

Generally, our paper give a brief review about metrics and methods of video quality assessment and introduce some basic concepts, methods and applications in this field. With the development of video acquisitions and applications, video quality assessment will play an more important role in the future. However, there are many challenges remaining to be resolved. It will be a long way to design new metrics or improve the old one to achieve high performance in the future. In the future, we planned to mine the quantitative relationship between video quality metrics and methods. This research can guide the development trend in the video quality assessment area. We will also follow the research on video quality improvement, which must have great need in the future.

**Acknowledgment** This research is supported by state grid corporation science and technology project “The pilot application on network access security for patrol data captured by unmanned planes and robots and intelligent recognition base on big data platform”.

## References

1. Agrawal R, Faloutsos C, Swami A (1993) Efficient similarity search in sequence databases
2. Alosioush AB, Sreelekha G (2011) Quality improvement of h. 264/avc frext by incorporating perceptual models. In: Proceedings of annual IEEE conference on india conference (INDICON), pp 1–6
3. Andrade H, Kurc T, Sussman A, Saltz J (2004) Optimizing the execution of multiple data analysis queries on parallel and distributed environments. *IEEE Trans Parallel Distrib Syst* 15(6):520–532
4. Anegekuh L, Sun L, Jammeh E, Is HM, Ifeachor E (2015) Content-based video quality prediction for hev encoded videos streamed over packet networks. *IEEE Trans Multimed* 17(8):1323–1334
5. Atenas M, Garcia M, Canovas A, Lloret J (2010) A mpeg-2/mpeg-4 quantizer to improve the video quality in iptv services. In: Proceedings of the 6th international conference on networking and services (ICNS), pp 49–54
6. Bhat A, Kannangara S, Zhao Y, Richardson I (2012) A full reference quality metric for compressed video based on mean squared error and video content. *IEEE Trans Circuits Syst Video Technol* 22(2):165–173
7. Boujut H, Benois-Pineau J, Ahmed T, Hadar O, Bonnet P (2011) A metric for no-reference video quality assessment for hd tv delivery based on saliency maps. In: IEEE international conference on multimedia and expo, pp 1–5
8. Chan SSM, Li Q, Wu Y, Zhuang Y (2002) Accommodating hybrid retrieval in a comprehensive video database management system. *IEEE Trans Multimed* 4(2):146–159
9. Chikkerur S, Sundaram V, Reisslein M, Karam LJ (2011) Objective video quality assessment methods: a classification, review, and performance comparison. *IEEE Trans Broadcast* 57(2):165–182
10. Choudary C, Liu T, Huang C-T (2007) Semantic retrieval of instructional videos. In: Proceedings of the 9th IEEE international symposium on multimedia workshops, pp 277–282
11. Dubravko U, Milan M, Vladimir Z, Maja P, Vladimir C, Dragan K (2011) Salient motion features for video quality assessment. *IEEE Trans Image Process* 20(4):948–958

12. Espina F, Morato D, Izal M, Magana E (2011) Improving video quality in network paths with bursty losses. In: Proceedings of IEEE conference on global telecommunications conference (GLOBECOM 2011), pp 1–6
13. Feng X, Liu T, Yang D, Wang Y (2008) Saliency based objective quality assessment of decoded video affected by packet losses. In: Proceedings of the 15th IEEE international conference on image processing (ICIP), pp 2560–2563
14. Feng X, Liu T, Yang D, Wang Y (2011) Saliency inspired full-reference quality metrics for packet-loss-impaired video. *IEEE Trans Broadcast* 57(1):81–88
15. Gunawan IP, Ghanbari M (2008) Reduced-reference video quality assessment using discriminative local harmonic strength with motion consideration. *IEEE Trans Circuits Syst Video Technol* 18(1):71–83
16. Ha K, Kim M (2011) A perceptual quality assessment metric using temporal complexity and disparity information for stereoscopic video. In: Proceedings of the 18th IEEE international conference on image processing (ICIP), pp 2525–2528
17. Hee M, Ik YY, Kim KC (1999) Hybrid video system supporting content-based retrieval. In: Proceedings of the 3rd international conference on computational intelligence and multimedia applications (ICCIMA), pp 258–262
18. Hewage CTER, Martini MG (2010) Reduced-reference quality evaluation for compressed depth maps associated with colour plus depth 3d video. In: Proceedings of the 17th IEEE international conference on image processing (ICIP), pp 4017–4020
19. Hirai K, Tumurtogoo J, Kikuchi A, Tsumura N, Nakaguchi T, Miyake Y (2010) Video quality assessment using spatio-velocity contrast sensitivity function. *IEICE Trans Inf Syst* 93(5):1253–1262
20. Jain A, Bhateja V (2011) A full-reference image quality metric for objective evaluation in spatial domain. In: International conference on communication and industrial application (ICCIA), pp 1–5
21. Jenkins C, Inman D (2000) Server-side automatic metadata generation using qualified dublin core and rdf. In: Proceedings of international conference on digital libraries: research and practice, pp 262–269
22. Kalpana S, Conrad BA (2010) Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Trans Image Process* 19(2):335–350
23. Kalpana S, Rajiv S, Conrad BA, Cormack LK (2010) Study of subjective and objective quality assessment of video. *IEEE Trans Image Process* 19(6):1427–1441
24. Karacali B, Krishnakumar AS (2012) Measuring video quality degradation using face detection. In: Proceedings of the 35th IEEE transactions on sarnoff symposium (SARNOFF), pp 1–5
25. Keimel C, Klimpke M, Habigt J, Diepold K (2011) No-reference video quality metric for hdtv based on h.264/avc bitstream features. In: IEEE international conference on image processing, pp 3325–3328
26. Kim D, Ryu S, Sohn K (2012) Depth perception and motion cue based 3d video quality assessment. In: Proceedings of IEEE international symposium on broadband multimedia systems and broadcasting (BMSB), pp 1–4
27. Kriegel H-P, Kroger P, Kunath P, Pryakhin A (2006) Effective similarity search in multimedia databases using multiple representations. In: Proceedings of the 12th international multi-media modelling conference proceedings, p 4
28. Kwok SH, Leon Zhao J (2006) Content-based object organization for efficient image retrieval in image databases. *Decis Support Syst* 42(3):1901–1916
29. Lee S-O, Jung K-S, Sim D-G (2010) Real-time objective quality assessment based on coding parameters extracted from h. 264/avc bitstream. *IEEE Trans Consum Electron* 56(2):1071–1078
30. Leszczuk M, Janowski L, Romaniak P, Papir Z (2013) Assessing quality of experience for high definition video streaming under diverse packet loss patterns. *Signal Process Image Commun* 28(8):903–916
31. Leszczuk M, Kowalczyk K, Janowski L, Papir Z (2015) Lightweight implementation of no-reference (nr) perceptual quality assessment of h. 264/avc compression. *Signal Process Image Commun* 39:457–465
32. Li C, Bovik AC (2010) Content-weighted video quality assessment using a three-component image model. *J Electron Imaging* 19(1):143–153
33. Li J, Xia Y, Shan Z, Liu Y (2015) Scalable constrained spectral clustering. *IEEE Trans Knowl Data Eng* 27(2):589–593
34. Ma Q, Zhang L, Wang B (2010) New strategy for image and video quality assessment. *J Electron Imaging* 19(1):011019–011019
35. Maalouf A, Larabi MC (2010) A no-reference color video quality metric based on a 3d multispectral wavelet transform. In: Proceedings of international workshop on quality of multimedia experience (QoMEX), pp 11–16
36. Martínez JM, Pereira F (2002) Mpeg-7: the generic multimedia content description standard, part 1. *IEEE Trans Multimed* 9(2):78–87

37. Matsubara FM, Hanada T, Imai S, Miura S, Akatsu S (2009) Managing a media server content directory in absence of reliable metadata. *IEEE Trans Consum Electron* 55(2):873–877
38. Narvekar ND, Karam LJ (2009) A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection. In: Proceedings of international workshop on quality of multimedia experience (QoMEX), pp 87–91
39. Narwaria M, Lin W, Liu A (2012) Low-complexity video quality assessment using temporal quality variations. *IEEE Trans Multimed* 14(3):525–535
40. Nezhadarya E, Ward RK (2013) Semi-blind quality estimation of compressed videos using digital watermarking. *Digital Signal Process* 23(5):1483–1495
41. Quan H-T, Ghanbari M (2010) Modelling of spatio-temporal interaction for video quality assessment. *Signal Process Image Commun* 25(7):535–546
42. Rasli RM, Haw S-C, Wong C-O (2010) A survey on optimizing video and audio query retrieval in multimedia databases. In: Proceedings of the 3rd international conference on advanced computer theory and engineering (ICACTE), vol 2, pp V2–302
43. Ribeiro C, David G, Calistru C (2004) A multimedia database workbench for content and context retrieval. In: Proceedings of the 6th IEEE workshop on multimedia signal processing, pp 430–433
44. Sadat ABMRI, Rubaiyat Islam BM, Lecca P (2009) On the performances in simulation of parallel databases: an overview on the most recent techniques for query optimization. In: Proceedings of international workshop on high performance computational systems biology (HIBI), pp 113–117
45. Seshadrinathan K, Bovik AC (2009) Motion-based perceptual quality assessment of video. In: IS&T/SPIE electronic imaging, pp 72400X–72400X
46. Shan Z, Xia Y, Hou P, He J (2016) Fusing incomplete multisensor heterogeneous data to estimate urban traffic. *IEEE MultiMedia* 23(3):56–63
47. Shen HT, Ooi BC, Zhou X (2005) Towards effective indexing for very large video sequence database. In: Proceedings of ACM SIGMOD international conference on management of data, pp 730–741
48. Shenoy ST, Ozsoyoglu ZM (1989) Design and implementation of a semantic query optimizer. *IEEE Trans Knowl Data Eng* 1(3):344–361
49. Steinacker A, Ghavam A, Steinmetz R (2001) Metadata standards for web-based resources. *IEEE Trans Multimed* 8(1):70–76
50. Tagliasacchi M, Valenzise G, Naccari M, Tubaro S (2010) A reduced-reference structural similarity approximation for videos corrupted by channel errors. *Multimed Tools Appl* 48(3):471–492
51. Torkamani-Azar F, Imani H, Fathollahian H (2015) Video quality measurement based on 3-d. singular value decomposition. *J Vis Commun Image Represent* 27:1–6
52. Vranješ M, Rimac-Drlje S, Grčić K (2013) Review of objective video quality metrics and performance comparison using different databases. *Image Commun* 28(1):1–19
53. Wang A, Jiang G, Wang X, Yu M (2009) New no-reference blocking artifacts metric based on human visual system. In: Proceedings of international conference on wireless communications & signal processing (WCSP), pp 1–5
54. Wang S-H, Chen W-L, Chiang T (2007) An efficient fgs to mpeg-1/2/4 single layer transcoder with r-d optimized multi-layer streaming technique for video quality improvement. *J Chin Inst Eng* 30(6):1059–1070
55. Wang Z, Lu L, Bovik AC (2004) Video quality assessment based on structural distortion measurement. *Signal Process Image Commun* 19(2):121–132
56. Weibel S (1997) The dublin core: a simple content description model for electronic resources. *Bull Am Soc Inf Sci Technol* 24(1):9–11
57. Wichterich M, Assent I, Kranen P, Seidl T (2008) Efficient emd-based similarity search in multimedia databases via flexible dimensionality reduction. In: Proceedings of ACM SIGMOD international conference on management of data, pp 199–212
58. Xia T, Mei T, Hua G, Zhang YD, Hua XS (2010) Visual quality assessment for web videos. *J Vis Commun Image Represent* 21(8):826–837
59. Xia Y, Chen J, Li J, Zhang Y (2016) Geometric discriminative features for aerial image retrieval in social media. *Multimed Syst* 22(4):497–507
60. Xia Y, Chen J, Lu X, Wang C, Xu C (2016) Big traffic data processing framework for intelligent monitoring and recording systems. *Neurocomputing* 181:139–146
61. Xia Y, Nie L, Zhang L, Yang Y, Hong R, Li X (2016) Weakly supervised multilabel clustering and its applications in computer vision. *IEEE Trans Cybern* 46(12):3220–3232
62. Xia Y, Chen W, Liu X, Zhang L, Li X, Xiang Y (2017) Adaptive multimedia data forwarding for privacy preservation in vehicular ad-hoc networks. *IEEE Trans Intell Transp Syst*
63. Xia Y, Liu Z, Yan Y, Chen Y, Zhang L, Zimmermann R (2017) Media quality assessment by perceptual gaze-shift patterns mining. *IEEE Trans Multimed*

64. Xia Y, Ren X, Peng Z, Zhang J, She L (2016) Effectively identifying the influential spreaders in large-scale social networks. *Multimed Tools Appl* 75(15):8829–8841
65. Xia Y, Shi X, Song G, Geng Q, Liu Y (2016) Towards improving quality of video-based vehicle counting method for traffic flow estimation. *Signal Process* 120:672–681
66. Xia Y, Zhang L, Hong R, Nie L, Yan Y, Shao L (2017) Perceptually guided photo retargeting. *IEEE Trans Cybern* 47(3):566–578
67. Xia Y, Zhang L, Tang S (2014) Large-scale aerial image categorization by multi-task discriminative topologies discovery. In: *Proceedings of the first international workshop on internet-scale multimedia management*, pp 53–58. ACM
68. Xia Y, Zhang L, Xu W, Shan Z, Liu Y (2015) Recognizing multi-view objects with occlusions using a deep architecture. *Inf Sci* 320:333–345
69. Xia Y, Zhu M, Gu Q, Zhang L, Li X (2016) Toward solving the steiner travelling salesman problem on urban road maps using the branch decomposition of graphs. *Inf Sci* 374:164–178
70. Xiang X, Shi Y, Guo L (2003) A conformance test suite of localized lom model. In: *Proceedings of the 3rd IEEE international conference on advanced learning technologies*, pp 288–289
71. Yamamura Y, Iwasaki S, Matsuo Y, Katto J (2013) Quality assessment of compressed video sequences having blocking artifacts by cepstrum analysis. In: *Proceedings of IEEE international conference on consumer electronics (ICCE)*, pp 494–495
72. Yao J, Xie Y, Tan J, Li Z, Qi J, Gao L (2012) No-reference video quality assessment using statistical features along temporal trajectory. *Procedia Eng* 29:947–951
73. Yoon J, Jayant N (2001) Relevance feedback for semantics based image retrieval. In: *Proceedings of international conference on image processing*, vol 1, pp 42–45
74. You J, Korhonen J, Perkiö A (2010) Spatial and temporal pooling of image quality metrics for perceptual video quality assessment on packet loss streams. In: *Proceedings of IEEE international conference on acoustics speech and signal processing (ICASSP)*, pp 1002–1005
75. You J, Korhonen J, Perkiö A, Ebrahimi T (2011) Balancing attended and global stimuli in perceived video quality assessment. *IEEE Trans Multimed* 13(6):1269–1285
76. Zeng K, Wang Z (2010) Temporal motion smoothness measurement for reduced-reference video quality assessment. In: *ICASSP*, pp 1010–1013
77. Zhang L, Gao Y, Ji R, Xia Y, Dai Q, Li X (2014) Actively learning human gaze shifting paths for semantics-aware photo cropping. *IEEE Trans Image Process* 23(5):2235–2245
78. Zhang L, Gao Y, Xia Y, Dai Q, Li X (2015) A fine-grained image categorization system by celllet-encoded spatial pyramid modeling. *IEEE Trans Ind Electron* 62(1):564–571
79. Zhang L, Gao Y, Xia Y, Lu K, Shen J, Ji R (2014) Representative discovery of structure cues for weakly-supervised image segmentation. *IEEE Trans Multimed* 16(2):470–479
80. Zhang L, Li X, Nie L, Yang Y, Xia Y (2016) Weakly supervised human fixations prediction. *IEEE Trans Cybern* 46(1):258–269
81. Zhang L, Liu X, Lu K (2014) Svd-based 3d image quality assessment by using depth information. In: *Proceedings of IEEE 17th international conference on computational science and engineering (CSE)*
82. Zhang L, Liu Z, Nie L, Li X et al (2016) Weakly-supervised multimodal kernel for categorizing aerial photographs. *IEEE Trans Image Process*
83. Zhang L, Xia Y, Ji R, Li X (2015) Spatial-aware object-level saliency prediction by learning graphlet hierarchies. *IEEE Trans Ind Electron* 62(2):1301–1308
84. Zhang L, Xia Y, Mao K, Ma H (2015) An effective video summarization framework toward handheld devices. *IEEE Trans Ind Electron* 62(2):1309–1316
85. Zhou W, Dao S, Kuo C-CJ (2002) On-line knowledge-and rule-based video classification system for video indexing and dissemination. *Inf Syst* 27(8):559–586



**Daojing He** received the B.Eng. (2007) and M. Eng. (2009) degrees from Harbin Institute of Technology (China) and the Ph.D. degree (2012) from Zhejiang University (China), all in Computer Science. He is currently a Professor in the Software Engineering Institute, East China Normal University, P.R. China. His research interests include network and systems security. He is an associate editor or on the editorial board of some international journals such as IEEE Communications Magazine and IEEE/KICS Journal of Communications and Networks.