Recent Advances and Challenges of Visual Signal Quality Assessment

MA Lin[1, 4], DENG Chenwei[2*], King Ngi Ngan[1], LIN Weisi[3]

[1]Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, China
[2]School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
[3]School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore
[4]Lenovo Corporate Research Hong Kong Branch, Hong Kong, China

Abstract: While quality assessment is essential for testing, optimizing, benchmarking, monitoring, and inspecting related systems and services, it also plays an essential role in the design of virtually all visual signal processing and communication algorithms, as well as various related decision-making processes. In this paper, we first provide an overview of recently derived quality assessment approaches for traditional visual signals (i.e., 2D images/videos), with highlights for new trends (such as machine learning approaches). On the other hand, with the ongoing development of devices and multimedia services, newly emerged visual signals (e.g., mobile/3D videos) are becoming more and more popular. This work focuses on recent progresses of quality metrics, which have been reviewed for the newly emerged forms of visual signals, which include scalable and mobile videos, High Dynamic Range (HDR) images, image segmentation results, 3D images/videos, and retargeted images.

Key words: objective quality assessment; 2D images and videos; human perception; newly emerged visual signals; Human Visual System

I. INTRODUCTION

The explosion in the number of computers and digital systems connected by networks such as the Internet has brought a flow of instant information into an increasingly large number of homes and businesses. Most of the information is in the form of digital visual signals (including image, video, etc.), as the most intuitive and faithful depiction of things in life and work. As a result, products (e.g., phone cameras) and services (e.g., YouTube) based upon visual signals have grown at an exponential rate.

Since human eyes are the ultimate receivers of the visual signals, how the quality of visual signals is gauged can be employed to dictate the formation of most of the corresponding technologies and applications. Quality measurement can be used as the benchmark of next-generation visual signal acquisition, compression, processing, transmission, and so on, to shape the full spectrum of technology development and enable new applications.

Perceptual quality of visual signals can be judged by human subjects in the most reliable way [1-2]. However, subjective assessment methods suffer from drawbacks as follows: 1) it is time-consuming, laborious and expensive, and requires many human subjects and repeated viewing sessions; 2) it is not feasible for on-line or real-time signal manipulations (such as encoding and transmission); 3) even in the cases where human assessment is possible (such as manufacturing assembly lines) and cost is not a problem, it depends upon the assessor's physical conditions, emotion states, personal experience, etc. Therefore, many researchers have carried out subjective tests to build

Received: 2013-02-19
Revised: 2013-03-28
Editor: Ming-Ting SUN

databases, which can help to efficiently evaluate the performances of the developed quality metrics. During the database construction, many observers are recruited into the subjective assessment. The subjective rating scores after processing are believed to be reliable to represent the true perceptual quality of the visual signal. The built databases include the representative image databases - LIVE, CSIQ, IVC, Toyama, TID, and A57, and the video databases - LIVE, EPFL, and VQEG FR-TV Phase I. Detailed information about these databases can be referred to Refs. [3-7].

Nowadays, automatic objective quality assessment has become an active research area, which aims at building computational models to predict perceptual signal quality. In other words, mathematical/engineering models are to be developed to take visual signals as inputs, while the output is a number to denote the perceived quality. Numerous objective metrics have been proposed for the prediction of multimedia quality. In Refs. [3-4], perceptual visual quality metrics are surveyed, which mostly focus on the quality assessment of traditional visual signals (i.e., 2D image/video). However, more and more visual signals are recently presented to the consumers in different formats, such as 3D image/video, mobile video, High Dynamic Range (HDR) images, and so on. Therefore, a systematic and up-to-date review of the visual signal quality assessment will be of great interest to the research community. In this paper, we aim at a comprehensive overview of the state-of-the-art research in the area of quality assessment of visual signals, especially for the newly emerged visual signals.

The rest of the paper is organized as follows. In Section II, some representative quality metrics for traditional visual signals are classified and reviewed, respectively, with highlights for the new trends (like machine learning approaches). In Section III, quality assessment of newly emerged visual signals will be reviewed in a more detailed way. Finally, Section IV will conclude this paper.

II. Quality Assessment for Traditional Visual Signals

Quality assessment methods can be categorized into different classes based on different criteria [3-6]. With regard to the availability of reference information, visual quality metrics for traditional visual signals can be generally classified into three board categories: Full-Reference (FR) metrics that need the complete reference signal, Reduced-Reference (RR) ones that need partial information of the reference signal, and Non-Reference (NR) ones that the processed signal is evaluated without any prior knowledge of the reference signal.

FR quality metric: Mean Square Error (MSE), Signal to Noise Ratio (SNR), and Peak SNR (PSNR) are the simplest and widely adopted FR metrics, where the pixel-wise signal fidelity is measured. Although they are appealing for optimization, they are demonstrated to be inconsistent with Human Visual System (HVS) perception, especially for the non-additive distortions, such as JPEG, and H.264/AVC compression. In order to handle the drawbacks of MSE and PSNR, in the past years, Perceptual Visual Quality Metrics (PVQMs) have been extensively investigated [8-21]. In Ref. [3], W. Lin and C.-C. Jay Kuo reviewed the progresses of the PVQMs during the past decades. The quality metrics can be roughly divided into model-based PVQMS, such as Refs. [11-12, 14-19] and signal-driven PVQMs, such as Refs. [8-10, 13]. After the review, two main trends of quality assessment are proposed for FR visual signal quality assessment, which are impairment decoupling [11, 22-27] and machine learning approaches [28-34], respectively.

For impairment decoupling approach, instead of treating the distortions indiscriminately, the distortions introduced into the visual signal are separated into different components. And different distortions are believed to correlate with HVS perception in different ways. In Refs. [11, 22-24], the authors separated the distortions into detail losses and additive impairments and treated them inde-

A comprehensive overview of recently derived quality metrics are reviewed for traditional visual signals and newly emerged visual signals, which include scalable and mobile videos, High Dynamic Range (HDR) images, image segmentation results, 3D images/videos, and retargeted images. The subjective evaluation and objective quality metrics are reviewed. It also discusses how to develop effective quality metrics for newly emerged visual signals.

pendently. In Ref. [25], the authors proposed to separate distortions into linear frequency distortions and additive noise degradations, which in essence can be interpreted as the two terms, i.e., detail losses and additive impairments. However, the decoupling algorithm in Ref. [25] was specifically designed for halftoning artifacts, while the works in Refs. [11, 22-24] are designed for general distortions. In Ref. [11], the HVS perception of detail losses is evaluated by calculating the Minkowski summations by referring to the restored and original image. For the additive impairments, the HVS responses can be further simulated by the Minkowski summation with the normalization by the number of pixels. In this way, the HVS responses to different impairments can be captured and finally combined together to indicate the corresponding perceptual quality.

The other new trend of PVQM design is machine learning approaches, the general framework of which is illustrated in Figure 1. The subjective databases provided the original and tested visual signals, as well as the corresponding subjective scores. Therefore, based on the provided information, the perceptual quality analysis process can be treated as a pattern recognition problem. Several useful features that are sensitive to visual signal perceptual quality are extracted from the original and tested visual signal, respectively. Based on the extracted features, machine learning algorithms will generate a trained model. Thereafter, the trained model will be employed to predict the perceptual quality. The machine

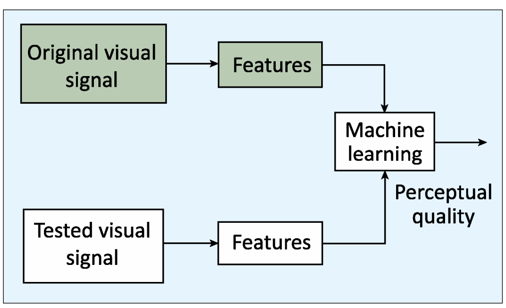

Fig.1 *General framework of machine learning approaches*

learning approaches are believed to have the abilities to learn complex data patterns and overcome the difficulties for modeling complex HVS properties.

In Ref. [30], the features of perceptual quality analysis are the singular vectors out of Singular Value Decomposition (SVD) to quantify major structural information. And the Support Vector Regression (SVR) is employed as the machine learning algorithm for automatic prediction of image quality. The use of SVR exploits the advantages of machine learning with the ability to learn complex data patterns for an effective and generalized mapping of features into a desired score, in contrast with the oft-utilized feature pooling process in the existing quality estimators. Also SVR can overcome the difficulty of model parameter determination for such a system to emulate the related, complex HVS modeling.

In Ref. [29], the image features are extracted based on Two-Dimensional (2D) mel-cepstrum, which is demonstrated that these features are effective since they can represent the structural information. In Ref. [33], the machine learning algorithm is employed to develop a Content-Dependent Multi-Metric Fusion (CD-MMF) for evaluating the image perceptual quality. A large number of image samples have been collected, which are associated with subjective scores and objective quality scores from different metrics. CD-MMF is set to be the nonlinear combination of multiple metrics with suitable weights obtained by a machine learning process. Moreover, the machine learning approach is employed to determine the context automatically for CD-MMF.

RR quality metric: RR quality metrics are roughly classified into three different but related categories. The first approach is based on image distortion modeling. The quality metrics [35-36] are mostly developed for the videos degraded by specific distortions, such as MPEG-2 compression. The second approach is developed on modeling HVS. The quality metrics [37-39] are developed by considering the HVS properties. For example, in Ref. [38], several HVS related features are extracted to indicate the spatial information losses, edge information changes, contrast information, and color impairments. Therefore, an effective RR quality metric named as VQM is developed, which has been adopted as a North American Standard by the American National Standards Institute (ANSI). The third type of approaches is based on modeling natural image/video statistics. The underline essential of these metrics [40-44] is that most real-world distortions will disturb the image statistics. The variations of the image statistics can be used to quantify the degradation level of the image/video. For example, in Ref. [41], Generalized Gaussian Density (GGD) is employed to depict the wavelet coefficient distribution. In Refs. [42, 44], GGD is employed to depict the coefficient distribution in Reorganized DCT (RDCT) domain, which can ensure better performances. As aforementioned, these three types of approaches are different but related. The developed RR metrics have either explicitly or implicitly considered these approaches together but to a limited extent. However, in order to develop a more effective RR quality metric, a more comprehensive consideration of these approaches is needed.

NR quality metric: NR quality metric design is an extremely difficult task [45]. Therefore, most of NR metrics focused on the image/video degraded by the specific distortions [46-49]. As JPEG 2000 employs the wavelet transform to compress the image, the wavelet statistical model is utilized to capture the compression distortion [50]. Liang et al. [46] combined the sharpness, blurring, and ringing measurements together to depict the perceptual quality of the JPEG 2000 coded image. The distribution of the DCT coefficient after quantization is modeled in Ref. [47] to predict the PSNR value of the JPEG coded image. Furthermore, Ferzli et al. [49] did the psychophysical experiment to test the blurring tolerance ability of the HVS, based on which the Just-Noticeable Blur (JNB) model is developed. These methods employ the behaviors of specific distortions to predict the degradation level. Therefore, if a new distortion is introduced, these methods can hardly evaluate the perceptual quality of the distorted image. In order to handle the drawbacks, generic NR metrics for images are researched [51-53], where different NR quality metrics designed for specific distortions are fused together without considering the reference visual signal. In Ref. [54], a comprehensive overview on the NR quality metric has been provided, which can be referred to for detailed information.

Tables I and II demonstrate the quality evaluation performances of the corresponding quality metrics (in LIVE image/video database). Three statistical parameters, specifically Linear Coefficient Correlation (LCC), Spearman-Rank Order Correlation Coefficient (SROCC), and Root Mean Square Error (RMSE) are employed to evaluate the metric's performance. It can be observed that PVQMs outperforms the simple PSNR, and FR PVQMs are generally more consistent with human perception, compared with those of RR and NR ones. FSIM [9] and MOVIE [19] perform

Table I *Comparison of different image quality metrics*

	Type	LCC	SROCC	RMSE
PSNR	FR	0.872	0.876	13.37
SSIM [8]	FR	0.904	0.910	11.68
VIF [13]	FR	0.960	0.963	7.673
FSIM [9]	FR	0.960	0.963	7.678
IGM [12]	FR	0.958	0.958	7.924
Machine learning approach [30]	FR	0.924	0.925	10.40
Impairment decoupling approach [11]	FR	0.936	0.946	9.627
WNISM [41]	RR	0.738	0.779	18.43
RR metric [42]	RR	0.883	0.879	12.84
NR metric [51]	NR	0.823	0.800	-
BLIINDS [52]	NR	0.680	0.663	20.01
DIIVINE [53]	NR	0.917	0.916	10.90

Table II *Comparison of different video quality metrics*

	Type	LCC	SROCC	RMSE
PSNR	FR	0.539 8	0.523 4	9.241
SSIM [8]	FR	0.499 9	0.524 7	9.507
MOVIE [19]	FR	0.811 6	0.789 0	-
VQM [38]	RR	0.716 0	0.702 9	7.664

the best for image and video quality assessment, respectively; but the machine learning and impairment decoupling approaches are quite promising and effective for evaluating visual signal perceptual quality.

III. Quality Assessment for Newly Emerged Visual Signals

Nowadays, newly emerged visual signals have issued new challenges for perceptual quality assessment, with the development of devices and multimedia services. In this section, we are to review recent research works on quality assessment of these visual signals, specifically scalable and mobile videos, HDR images, image segmentation results, 3D images/videos, and retargeted images.

3.1 Scalable and mobile video quality assessment

Nowadays, with development of devices, such as smartphones and tablets, High Definition (HD) video sequences need to be displayed in different resolutions. Therefore, scalable video coding is needed to meet the requirements. Therefore, Video Coding Experts Group (VCEG) released a call for proposals on Scalable Video Coding (SVC) extensions of High Efficient Video Coding (HEVC) [55]. Also the perceptual quality assessment of SVC is very important, as the human eyes are the final receivers of the distributed video sequences.

In Ref. [56], the authors have performed a large-scale study on the subjective quality of the SVC for content distribution. The influence of the combination of scalable parameters in SVC was studied. The subjective results of SVC are analyzed with respect to five dimensions, namely, video codec, content, spatial resolution, temporal resolution, and frame quality. Based on the study, guidelines are provided for an adaptation strategy of SVC that can select the optimal scalability options for resource-constrained networks. The adaptation strategy can be summarized as follows. First, when the bit-rate is small and only layers having small spatial resolutions are available, a larger spatial resolution is preferable in order to obtain the lowest acceptable frame quality without strong blurring. In the case of H.264/SVC, this observation was valid even when the frame rate decreases along with increase of the spatial resolution. Second, for large bit-rate conditions, acceptable frame quality is achieved and thus a high frame rate, which is obtained at the cost of the decreased pixel bit-rate, becomes important for better subjective quality. The built database has been made publicly available to the research community working on objective quality design to measure the video sequences generated by SVC [56]. The database was created by a pair-wise comparison between two stimuli, i.e., SVC video sequences. The participated subject was asked to indicate which one had better quality. In total, sixteen subjects (11 men and 5 women) participated in the subjective evaluation procedure. Also in Ref. [57], video perceptual quality was analyzed in the sense of considering multiple dimensions, specifically the encoder type, video content, bit rate, frame size, and frame rate. It was reported that the video perceptual quality is affected by the encoder type, video content, bit rate, frame rate, and frame size in a descending order of significance. These findings are very useful for cross-dimensional video assessment and video adaptation for video distribution.

Moreover, with development of the widely spread and publicly adopted smartphones and tablets, mobile videos are more and more useful and important for mobile communications and information sharing. Therefore, the perceptual qualities of mobile videos are very important. In Ref. [58], the authors built a new subjective mobile video database, where the human study was performed on mobile phones and tablets in order to gauge the human perception of qualities on mobile devices. The LIVE mobile database includes distortions that have been studied such as compression and wireless packet-loss, which also incorporates dynamically varying distortions that change as a function of time, such as frame-freezes and temporally varying compression rates. The subjective study portion of the database includes both the DMOS computed from the ratings that the subjects provided at the end of each video clip, as well as the continuous temporal scores that the subjects recorded as they viewed the video. The study involved over 50 subjects and resulted in 5 300 summary subjective scores and time-sampled subjective traces of quality. The human opinion is analyzed using statistical techniques, and also studied by a variety of models of temporal pooling that may reflect strategies that the subjects used to make the final decision on video quality. Also quality ratings obtained from the tablet and the mobile phone studies were to study the impact of these different display modes on quality.

Several objective image and video quality assessment algorithms with regards to their efficacy in predicting visual quality are evaluated over the database, which is illustrated in Table III to provide the benchmark results. And the LIVE Mobile VQA database, along with the subject DMOS and the continuous temporal scores is being made available to researchers in the field of VQA at no cost in order to further research in the area of video quality assessment. It can be observed that most of the work of scalable and mobile video quality assessment focuses on subjective testing method to build subjective database. The traditional quality metrics, such as PSNR and SSIM are employed to evaluate the perceptual quality. However, they are demonstrated to be inefficient. Therefore, in order to develop an effective quality metric for scalable and mobile videos, the video content, bit rate, frame rate, frame rate, etc. should be considered.

3.2 HDR images quality assessment

HDR images are becoming more widely available, due to recent advances in imaging and computer graphics technologies. Tone Mapping Operators (TMOs) are employed to visualize HDR images on standard display devices. How to convert HDR images by tone mapping operators and how to evaluate the tone mapped images are now researched in Refs. [59-61].

Table III *Quality metric comparisons on the mobile database*

	Tablet			Mobile		
	SROCC	LCC	RMSE	SROCC	LCC	RMSE
PSNR	0.588 6	0.634 8	0.663 0	0.678 0	0.690 9	0.667 0
SSIM [8]	0.430 0	0.489 3	0.748 3	0.649 8	0.663 7	0.690 1
VIF [13]	0.726 1	0.763 5	0.554 1	0.743 9	0.787 0	0.569 2
VQM [38]	0.555 2	0.615 0	0.698 0	0.694 5	0.702 3	0.666 3
MOVIE [19]	0.679 2	0.782 8	0.534 2	0.642 0	0.715 7	0.644 4

Due to the reduction in dynamic range, TMOs cannot preserve all information in HDR images, and human observers of the LDR versions of these images may not be aware of this. It is demonstrated that structural fidelity plays an important role in Refs. [59-60]. However, structural fidelity alone does not suffice to provide an overall quality evaluation. A good quality tone mapped image should achieve a good tradeoff between structural fidelity preservation and statistical naturalness. The overall structural fidelity is defined according to:

$$S = \prod_{l=1}^{L} S_l^{\beta_l} \tag{1}$$

where L is the total number of scales and β_l is the weight assigned to the l-th scale. And the single score S_l is calculated by:

$$S_l = \frac{1}{N_l} \sum_{i=1}^{N_l} S_{\text{local}}(x_i, y_i) \tag{2}$$

where x_i and y_i are the i-th patches in the HDR and LDR images being compared, respectively, and N_l is the number of patches in the l-th scale. The statistical naturalness model is built upon statistics conducted on about 3 000 8-bits/pixel gray-scale images, which can represent many different types of natural scenes. The means and standard deviations of these images can reflect the global intensity and contrast of images. And the histograms of mean and standard deviation can be well fitted using a Gaussian P_m and a Beta probability density function P_d. Recent studies suggested that brightness and contrast are largely independent quantities in terms of natural image statistics and biological computation. As a result, their joint probability density function would be the product of two. The statistical naturalness measure can be defined as:

$$N = \frac{1}{k} P_m P_d \tag{3}$$

where k is the normalization factor given by $k = \max\{P_m, P_d\}$. This constrains the statistical naturalness measure to be bounded between 0 and 1.Finally, the quality model can be determined by:

$$Q = \alpha S^{\alpha} + (1-\alpha) N^{\beta} \tag{4}$$

In Ref. [62], an overview about the effect of the basic image attributes in the HDR tone mapping is presented. Results of subjective psychophysical experiments that the authors performed are publicly available to researchers. Moreover, the evaluations of existing TMOs are provided based on the HDR database. With validation on the subject-rated image database [62], the metric is demonstrated to be effective as shown in Table IV. The structural fidelity and image naturalness are both correlated to HVS perception of HDR images. Furthermore, the quality metric can be further applied to tone mapping operators by parameter tuning and adaptive fusion of multiple tone mapped images.

In Ref. [61], a novel image quality metric is proposed to be capable of operating on an image pair with arbitrary dynamic ranges, which defines a new HVS model based on the detection and classification of visible changes in the image structure. The metric can be considered as a hybrid of contrast detection and structural similarity metrics. However, instead of detecting contrast changes, the proposed metric is sensitive to three types of structural changes. 1) Loss of visible contrast happens when a contrast that was visible in the reference image becomes invisible in the test image. This typically happens when a TMO compresses details to the level that they become invisible. 2) Amplification of invisible contrast happens when a contrast that was invisible in the reference image becomes visible in the test image. For instance, this can happen when contouring artifacts starts to appear due to contrast stretching in the inverse TMO application. 3) Reversal of visible contrast happens when a contrast is visible in both reference and test images, but has different polarity. This can be observed at image locations with strong distortions, such as clipping or salient compression artifacts.

Table IV *Performance of HDR quality metric* [59-60]

	Structural fidelity	Naturalness	Overall
KROCC	0.692 3	0.384 6	0.717 9
SROCC	0.791 2	0.538 5	0.818 7

It should be noted that the TMOs should try their best to preserve the structural information in order to visualize HDR images on standard display devices. Therefore, the quality metrics [59-61] try to evaluate the structural distortion between high dynamic and low dynamic images. Moreover, after TMO processing, naturalness of the images is also considered in Refs. [59-60], which further affects perceptual quality.

3.3 Image segmentation quality assessment

Image segmentation is a fundamental problem in computer vision. Over the past decades, a large number of segmentation algorithms have been proposed, which makes the evaluation of perceptual correctness on the segmentation output becomes a demanding task. The existing objective evaluation methods for segmentation can be classified as ground-truth based ones and none-ground-truth based ones.

In non-ground-truth based methods, the empirical goodness measures are proposed to meet the heuristic criteria in the desirable segmentations. The quality score is calculated based on these criteria to predict the segmentation quality. There have been examples of empirical measures on the uniformity of colors [63-64] or luminance [65] and even the shape of object regions [66]. Since the criteria are mainly summarized from the common characteristics or semantic information of the objects (e.g., homogeneous regions, smooth boundaries, etc.), they are not accurate enough to describe the complex objects in the images.

The ground-truth based methods measure the difference between the segmentation result and the human-labeled ground truths. They are more intuitive than the empirical based measures, since the ground truths can well represent the human-level interpretation of an image. Some measures in this category aim to count the degree of overlapping between regions with strategies of being intolerant [67] or tolerant [68] to region refinement. In contrast to working on regions, there are also measures [69-70] matching the boundaries between segmentations. Considering only the region boundaries, these measures are more sensitive to the dissimilarity between the segmentation and the ground truths than the region based measures. Some other measures use non-parametric tests to count the pairs of pixels that belong to the same region in different segmentations. The well-known Rand index [71] and its variants [72-73] are of this kind.

In Ref. [74], the authors provided a framework to quantitatively evaluate the quality of a given segmentation with multiple ground truth segmentations. It is assumed that if a segmentation is "good", it can be constructed by pieces of the ground truth segmentations. And a new ground truth is adaptively constructed, which can be locally matched to the segmentation as much as possible and preserve the structural consistency. The quality of the segmentation can then be evaluated by measuring its distance to the adaptively composite ground truth. There is a public segmentation database named as the Berkeley Segmentation Dataset and Benchmark. The goal of this database is to provide an empirical basis for research on image segmentation and boundary detection. To this end, 12 000 hand-labeled segmentations of 1 000 Corel dataset images from 30 human subjects are collected. Half of the segmentations were obtained from presenting the subject with a color image; the other half from presenting a grayscale image. The public benchmark based on this data consists of all of the grayscale and color segmentations for 300 images. The images are divided into a training set of 200 images, and a test set of 100 images. With evaluation on the benchmark Berkeley database, the method can faithfully reflect the perceptual quality of the segmentation. During the evaluation process, the commonly used F-test [75] and the probabilistic rand index [76] are computed on a large amount of segmentation results.

Table V shows the comparison results of the three measures. Each of them gives 500 results on the pair of segmentations, where our measure has 379 results which are consistent to the human judgment, while the F-measure and the probabilistic rand index have only 333 and 266, respectively. The number of results which are only correctly produced by one measure is named as the "winning" case. The proposed method [74] obtains 114 winning cases, while F-measure and probabilistic rand index only obtain 4 and 19, respectively. And there are 16 results which are wrongly classified by all of the three measures. The proposed measure outperforms the other two in both of the "consistent" and the "winning" cases. However, so far there is no standard procedure for segmentation quality evaluation due to the ill-defined nature of image segmentation, i.e., there might be multiple acceptable segmentations which are consistent to the human interpretation of an image. In addition, there exists a large diversity in the perceptually meaningful segmentations for different images. The above factors make the evaluation very complex.

3.4 3D images and videos

Stereoscopic/3D Image and Video Quality Assessment (IQA/VQA) has become increasing relevant in today's world, owing to the amount of attention that has recently been focused on 3D/stereoscopic cinema, television, gaming, and mobile video. With the development of the devices, more and more 3D visual signals are provided for the consumers, which can generate better Quality of Experience (QoE). However,

Table V *The number of results which are consistent to the human judgment, as well as the numbers of "winning" cases and failure cases by competing methods*

	Consistent	Winning	Failure
F-test [75]	333	4	16
Index in Ref. [76]	266	19	
Method in Ref. [74]	379	114	

understanding the quality of experience of human viewers as they watch 3D videos is a complex and multi-disciplinary problem. Therefore, many researchers have devoted the efforts for creating benchmark dataset and quality metrics of 3D visual signals [77-80].

In Ref. [77], LIVE 3D IQA database incorporates "true" depth information along with stereoscopic pairs and human opinion scores. The LIVE 3D IQA database consists of 20

Table VI *Performance of 2D IQA algorithms in predicting perceived 3D image quality*

		JP2K	JPEG	White Gaussian noise	Blur	Fast-fading	All
PSNR	SROCC	0.796 7	0.131 1	0.931 8	0.901 6	0.595 7	0.837 0
	LCC	0.788 9	0.231 1	0.934 7	0.893 7	0.706 2	0.825 1
	RMSE	7.958 7	6.362 4	5.914 5	6.527 1	8.797 1	9.267 8
SSIM	SROCC	0.857 2	0.434 6	0.939 5	0.882 2	0.584 9	0.877 2
	LCC	0.865 0	0.484 9	0.937 4	0.919 7	0.721 2	0.872 7
	RMSE	6.498 4	5.719 1	5.794 7	5.681 4	8.606 9	8.005 9
VIF	SROCC	0.901 8	0.582 8	0.932 5	0.931 2	0.803 7	0.920 4
	LCC	0.936 1	0.673 8	0.927 3	0.957 0	0.854 2	0.918 3
	RMSE	4.557 0	4.831 9	6.229 1	4.198 6	6.461 5	6.490 3
BIQI[81]	SROCC	0.772 7	0.488 7	0.927 7	0.859 6	0.706 7	0.865 2
	LCC	0.820 3	0.613 6	0.932 3	0.899 5	0.776 2	0.879 2
	RMSE	7.410 0	5.163 6	6.016 6	6.323 8	7.867 9	7.811 9

Table VII *Performance of 3D IQA algorithms in predicting perceived 3D image quality*

		JP2K	JPEG	White Gaussian noise	Blur	Fast-fading	All
Benoit [82]	SROCC	0.910 3	0.602 8	0.929 2	0.930 8	0.698 9	0.899 2
	LCC	0.939 8	0.640 5	0.925 3	0.948 8	0.747 2	0.902 5
	RMSE	4.426 6	5.022 0	6.307 6	4.571 4	8.257 8	7.061 7
Hewage[83]	SROCC	0.855 8	0.500 1	0.896 3	0.690 0	0.544 7	0.814 0
	LCC	0.904 3	0.530 5	0.895 5	0.798 4	0.669 8	0.830 3
	RMSE	5.530 0	5.543 1	7.405 6	8.748 0	9.226 3	9.139 3
You [84]	SROCC	0.859 8	0.438 8	0.939 5	0.882 2	0.588 3	0.878 9
	LCC	0.877 8	0.487 4	0.941 2	0.919 8	0.730 0	0.881 4
	RMSE	6.206 6	5.709 7	5.621 6	5.679 8	8.492 3	7.746 3
Gorley [85]	SROCC	0.420 3	0.015 2	0.740 8	0.749 8	0.366 3	0.141 9
	LCC	0.485 3	0.312 4	0.796 1	0.852 7	0.364 8	0.451 1
	RMSE	11.323 7	6.211 9	10.197 9	7.562 2	11.569 1	14.635 0
Shen [86]	SROCC	0.213 3	0.244 0	0.891 7	0.658 6	0.266 5	0.067 9
	LCC	0.503 9	0.389 9	0.898 8	0.684 6	0.483 0	0.574 3
	RMSE	12.275 4	6.021 6	7.293 9	10.554 7	10.882 0	13.547 3
Yang [87]	SROCC	0.150 1	0.132 8	0.847 1	0.326 6	0.142 6	0.078 5
	LCC	0.201 2	0.273 8	0.870 1	0.626 1	0.282 4	0.390 9
	RMSE	12.697 9	6.289 4	8.200 2	12.129 1	11.946 2	15.248 1
Zhu [88]	SROCC	0.770 8	0.292 9	0.465 1	0.793 5	0.475 2	0.638 8
	LCC	0.807 3	0.379 0	0.517 8	0.777 0	0.503 8	0.626 3
	RMSE	7.681 3	6.068 4	14.720 1	9.127 0	10.736 2	12.782 8
Akhter [89]	SROCC	0.865 7	0.675 4	0.913 7	0.554 9	0.639 3	0.382 7
	LCC	0.905 9	0.729 4	0.904 7	0.617 7	0.660 3	0.427 0
	RMSE	5.483 6	4.473 6	7.092 9	11.387 2	9.332 1	14.827 4

reference images, 5 distortion categories and a total of 365 distorted images along with the associated DMOS. The authors described the creation of the database and analyzed the performance of a variety of 2D and 3D quality models using the new database. The database as well as the algorithms evaluated is available for researchers in the field to use in order to enable objective comparisons of future algorithms. Also the authors broadly summarized the field of 3D QA focusing on key unresolved problems including stereoscopic distortions, 3D masking, and algorithm development. The traditional 2D image quality metrics are employed to evaluate the 3D stereoscopic image qualities, which is illustrated in Table VI. Also several 3D image quality metrics are employed to provide the benchmark performances shown in Table VII, such as: Benoit [82], Hewage [83], You [84], Gorley [85], Shen [86], Yang [87], Zhu [88], and Akhter [89].

Based on the results in Table VI, it is clear that for the set of distortions considered, the 2D IQA algorithms perform well in terms of correlation with human subjectivity, while the addition of disparity/depth in the 3D algorithms cannot materially improve the performance. However, disparity activity (e.g., caused by rapid changes in depth) may affect distortion visibility, which can be interpreted from the human experiences. Also it can be observed from Table VII that the 3D metric Benoit [82] can generate the best performance over the built database. However, the performance is still not good enough. The reason is that almost of the 3D QA algorithms are simple extensions of 2D QA algorithms with some additional "features" extracted from depth (generally disparity differences). The way in which this disparity information is incorporated into these 3D QA algorithms is not based on any perceptual principles.

As we all know, the depth information is very important for the 3D video application, such as displaying, rendering, and so on. In Ref. [78], the authors created both a subjective experiment and a new objective quality metric to evaluate the depth perception of 3D stereoscopic videos. Moreover, the authors in Ref. [79] investigated the Depth Image-Based Rendering (DIBR) based synthesized view evaluation problem. Different view synthesis algorithms are evaluated through subjective and objective measurements. The hints for developing new objective quality metric for 3D synthesized view are discussed. The authors in Ref. [80] derived the RR quality metric for 3-D videos. The information of depth edges and color images in the areas in the proximity of edges are extracted for the RR quality metric, which can be utilized for 3-D video compression and transmission.

Overall, the field of 3D QA remains an extremely interesting one. There is tremendous scope for research in this area. The large gaps in our understanding of the perception of stereoscopic distortions and of appropriate statistical models for 3D natural scenes are still need to be researched. A multi-pronged approach combining research and concepts from the visual sciences and image processing will hopefully lead to models that predict stereoscopic quality with high accuracy.

3.5 Retargeted images

The image retargeting methods are proposed to adjust the source images into arbitrary sizes for displaying. Therefore, there issues a new challenge of objectively evaluating the retargeted image perceptual quality, where the resolution has been changed, the objective shape may be distorted, and some content information may be discarded. Till now, there are two main public images retargeting quality databases [90-92]. In Ref. [90], the database is built by concentrating on a comparative study of existing retargeting methods. The authors compared which retargeting method generates the retargeted image with the highest perceptual quality. The subjective test is performed in a pair comparison way, where the participants are shown two retargeted images at a time, side by side, and are asked to simply choose the one they like better. The resulting database comprises the retargeted image and the corresponding number of times that the retargeted image is favored over another one.

In Refs. [91-92], a subjective study is conducted to assess the perceptual quality of the retargeted image to build a publicly available database. Totally 171 retargeted images (in two different scales) are generated by different retargeting methods from 57 source images. With the source image as the reference, the perceptual quality of each retargeted image has been subjectively rated by at least 30 human viewers on a pre-defined scale. After processing the subjective ratings, the MOS value and the corresponding standard deviation are obtained for each image. Based on the MOS values, the built image retargeting database was analyzed from the perspectives of the retargeting scale, the retargeting method, and the source image content. Moreover, some publicly available quality metrics for retargeted images are evaluated on the built database.

There are several existing quality metrics on retargeted images, such as Edge Histogram (EH) [93], SIFT-flow [94], Bidirectional Similarity (BDS) [95-96], and Earth's Mover Distance (EMD) [97-98]. EH captures the spatial distribution of edges in the image. In order to depict the local edge distribution, the image is divided into 4×4 sub-images, each of which is examined by 5 different orientations: vertical, horizontal, two diagonals, and isotropic (non-directional). For each sub-image, a normalized 5-bin histogram is obtained by classifying apparent edges to these five categories. The feature is defined to be the combination of these histograms, which results in 4×4×5 = 80 length description. Only the intensity component is employed for edge detection. And the -norm distance is employed to measure the feature distance between two images. SIFT-flow descriptors characterize view-invariant and brightness-independent image structures. Matching SIFT descriptors allows establishing meaningful correspondences across image with significantly different image content. Furthermore, the pixel displacement (indicating by the SIFT correspondence matching) should be spatial coherent, which means that close-by pixels should have similar displacement. BDS tries to capture that two signals (original and retargeted images) are considered to be "visually similar" is as many as possible patches of one image (at multiple scales) are contained in another image, and vice versa. EMD is based on the minimal cost that must be paid to transform one distribution into the other. The signature or histogram, which represents a set of feature clusters, is viewed as the histogram distribution. The point is the central value in bin of the histogram, and is to indicate the corresponding proportion. The definition of cluster is open. The color, position, and texture information can be employed to obtain the feature clusters. Only the size of the clusters in the feature space needs to be limited. EMD is defined as the minimum work normalized by the total flow to convert one histogram to another.

The performances evaluated on the database are shown in Table VIII. It can be observed that all of the existing metrics perform poorly on the database. For the EMD, the composed histogram only represents the feature distribution of the image, which cannot accurately depict the object shape and the content information of the image. Therefore, the shape distortions and content information loss, introduced during the retargeting process, are not effectively described. BDS tries to capture how much information one image conveys of the other image in a bidirectional way. However, although it is claimed that the spatial geometric relationship is considered by a multiple scale approach, the order-relationship can still not be preserved, such as the local-order of each pixel or patch. Therefore, the dissimilarity metric of BDS does not accurately depict the object shape distortion either. SIFT-flow employs the SIFT descriptor to detect the correspondence between two images. It is claimed that the order-relationship

Table VIII *Metric performance on image retargeting database*

	EH	EMD	BSE	SIFT-flow
LCC	0.342 2	0.276 0	0.289 6	0.314 1
SROCC	0.328 8	0.290 4	0.288 7	0.289 9
RMSE	12.686	12.977	12.922	12.817

of the pixels or patches is captured. However, the content information loss during the retargeting process is not considered. EH employs the edge histograms to describe the image, which are organized in order for comparison. EH can somehow represent the object shape information in the image. Same as the SIFT- flow, the content information loss is not accounted.

What have been described above are the reasons why the metrics available cannot perform effectively for retargeted images. It can be observed that most of the quality metrics try to extract several features from the original and retargeted images, and compare the differences for quality assessment. However, these features cannot accurately capture the shape distortions and content information loss, which have been introduced into the retargeted images. Therefore, in order to develop a more effective quality metric for retargeted images, the shape distortions and content information loss should be first accurately captured and then fused together, as a suggested direction of future research.

IV. Conclusion

In this paper, an overview and comparison of objective quality metrics for visual signals was provided, for both traditional and newly emerged visual signals, as well as giving comments and discussion whenever appropriate. Specifically, we highlighted the new trends of traditional visual signal quality assessment; those related to newly emerged visual signals have been discussed in a more detailed way, and such new forms of visual signals being discussed include scalable and mobile video, HDR image, image segmentation, 3D images/video, and retargeted images. We have also shared our views in some future development in the areas.

Acknowledgement

The authors would like to thank the reviewers for their detailed reviews and constructive comments, which have helped to improve the quality of this paper. The work described in this paper was partially supported by the Research Grants Council of the Hong Kong SAR, China (Project CUHK 415712); and the Ministry of Education Academic Research Fund (AcRF) Tier 2 in Singapore under Grant No. T208B1218.

References

[1] ITU-T-P.800. Methods for Subjective Determination of Transmission Quality[S], 1996.

[2] ITU-R BT.500-11. Methodology for the Subjective Assessment of the Quality of Television Pictures[S], 2000.

[3] LIN Weisi, KUO J C. Perceptual Visual Quality Metrics: A Survey[J]. Journal of Visual Communication and Image Representation, 2011, 22(4): 297-312.

[4] LIU T J, LIN Weisi, KUO J C.-C. Recent Developments and Future Trends in Visual Quality Assessment[C]// Proceedings of Asia-Pacific Signal and Information Processing Association Annual Submit and Conference, 2011 (APSIPA ASC 2011): October 18-21, 2011. Xi'an, China, 2011.

[5] ENGELKE U, ZEPERNICK H. Perceptual-Based Quality Metrics for Image and Video Services: A Survey[C]// Proceedings of the 3rd EuroNGI Conference on Next Generation Internet Networks: May 21-23, 2007. Trondheim, Norway, 2007: 190-197.

[6] WINKLER S, MOHANDAS P. The Evolution of Video Quality Measurement: from PSNR to Hybrid Metrics[J]. IEEE Transactions on Broadcasting, 2008, 54(3): 660-668.

[7] 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.

[8] WANG Zhou, BOVIK A C, SHERKH H R, *et al.* Image Quality Assessment: from Error Visibility to Structural Similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[9] MA Lin, LI Songnan, NGAN K N. Visual Horizontal Effect for Image Quality Assessment[J]. IEEE Signal Processing Letters, 2010, 17(7): 627-630.

[10] LIU Anming, LIN Weisi, NARWARIA M. Image Quality Assessment Based on Gradient Similarity[J]. IEEE Transactions on Image Processing, 2012, 21(4): 1500-1512.

[11] LI Songnan, ZHANG Fan, MA Lin, *et al.* Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments[J]. IEEE Transactions on Multimedia, 2011, 13(5): 935-949.

[12] WU Jinjian, LIN Weisi, SHI Guangming, *et al*. A Perceptual Quality Metric with Internal Generative Mechanism[J]. IEEE Transactions on Image Processing, 2013, 22(1): 43-54.
[13] SHEIKH H R, BOVIK A C. Image Information and Visual Quality[J]. IEEE Transactions on Image Processing, 2006, 15(2): 430-444.
[14] DALY S. The Visible Difference Predictor: An Algorithm for the Assessment of Image Fidelity[M]. Cambridge, CA, USA: Digital Images and Human Vision, MIT Press, 1993.
[15] WINKLER S. Digital Video Quality: Vision Models and Metrics, John Wiley and Sons[M]. Portland: Book News Inc., 2005.
[16] ITU-T Recommendation J.247. Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference[S], 2008.
[17] YANG X K, LING W S, LU Z K, *et al*. Just Noticeable Distortion Model and Its Applications in Video Coding[J]. Signal Processing: Image Communication, 2005, 20(7): 662-680.
[18] MA Lin, NGAN K N, ZHANG Fan, *et al*. Adaptive Block-Size Transform Based Just-Noticeable Difference Model for Images/Videos[J]. Signal Processing: Image Communication, 2011, 26(3): 162-174.
[19] 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.
[20] ZHANG Fan, MA Lin, LI Songnan, *et al*. Practical Image Quality Metric Applied to Image Coding[J]. IEEE Transactions on Multimedia, 2011, 13(4): 615-624.
[21] NARWARIA M, LIN Weisi. SVD-Based Quality Metric for Image and Video Using Machine Learning[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2012, 42(2): 347-364.
[22] LI Songnan, MA Lin, NGAN K N. Full-Reference Video Quality Assessment by Decoupling Detail Losses and Additive Impairments[J]. IEEE Transactions on Circuit System for Video Technology, 2012, 22(7): 1100-1112.
[23] LI Songnan, MA Lin, NGAN K N. Video Quality Assessment by Decoupling Additive Impairments and Detail Losses[C]// Proceedings of the 3rd International Workshop on Quality of Multimedia Experience (QoMEX 2011): September 7-9, 2011. Mechelen, Belgium, 2011: 90-95.
[24] LI Songnan, NGAN K N. Subtractive Impairment, Additive Impairment, and Image Visual Quality[C]// Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS 2010): May 30-June 2, 2010. Paris, France, 2010: 3373- 3376.
[25] DAMERA-VENKATA N, KITE T D, GEISLER W S, *et al*. Image Quality Assessment Based on a Degradation Model[J]. IEEE Transactions on Image Processing, 2000, 9(4): 636-650.
[26] LARSON E C, CHANDLER D M. Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy[J]. Electronic Imaging, 2010, 19(1): 1-21.
[27] ROUSE D M, PEPION R, HEMAMI S S, *et al*. Image Utility Assessment and a Relationship with Image Quality Assessment[J]. Human Vision Electronic Imaging, 2009, 7240: 724010-1-724010-14.
[28] NARWARIA M, LIN Weisi. SVD-Based Quality Metric for Image and Video Using Machine Learning[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 2012, 42(2): 347-364.
[29] NARWARIA M, LIN Weisi, CETIN A E. Scalable Image Quality Assessment with 2D Mel-Cepstrum and Machine Learning Approach[J]. Pattern Recognition, 2011, 45(1): 299-313.
[30] NARWARIA M, LIN Weisi. Objective Image Quality Assessment Based on Support Vector Regression[J]. IEEE Transactions on Neural Networks, 2010, 21(3): 515-519.
[31] LUO Huitao. A Training-Based No-Reference Image Quality Assessment Algorithm[C]// Proceedings of 2004 International Conference on Image Proceeding (ICIP 2004): October 24-27, 2004. Singapore, 2004: 2973-2976.
[32] SURESH S, BABU V, SUNDARARAJAN N. Image Quality Measurement Using Sparse Extreme Learning Machine Classifier[C]// Proceedings of the 9th International Conference on Control, Automation, Robotics and Vision, 2006 (ICARCV 2006): December 5-8, 2006. Singapore, 2006: 1-6.
[33] LIU T J, LIN Weisi, KUO C.-C J. A Multi-Metric Fusion Approach to Visual Quality Assessment[C]// Proceedings of 2011 the 3rd International Workshop on Quality of Multimedia Experience (QoMEX): September 7-9, 2011. Mechelen, Belgium, 2011: 72-77.
[34] NARWARIA M, LIN Weisi. Video Quality Assessment Using Temporal Quality Variations and Machine Learning[C]// Proceedings of 2011 IEEE International Conference on Multimedia and Expo (ICME): July 11-15, 2011. Barcelona, Spain, 2011: 1-6.
[35] WOLF S, PINSON M. Low Bandwidth Reduced Reference Video Quality Monitoring System [C]// Proceedings of the 1st International Workshop on Video Processing and Quality Metrics

for Consumer Electronics: January, 2005.

[36] TAGLIASACCHI M, VALENZISE G, NACCARI M, *et al*. A Reduced-Reference Structural Similarity Approximation for Videos Corrupted by Channel Errors[J]. Multimedia Tools Application, 2010, 48(3): 471-492.

[37] LE CALLET P, VIARD-GAUDIN C, BARBA D. A Convolutional Neural Network Approach for Objective Video Quality Assessment[J]. IEEE Transactions on Neural Network, 2006, 17(5): 1316-1327.

[38] PINSON M.H, WOLF S. A New Standardized Method for Objectively Measuring Video Quality[J]. IEEE Transactions on Broadcasting, 2004, 50(3): 312-322.

[39] REDI J.A, GASTALDO P, HEYNDERICKX I, *et al*. Color Distribution Information for Reduced-Reference Assessment of Perceived Image Quality[J]. IEEE Transactions on Circuit System Video Technology, 2010, 20(12): 1757-1769.

[40] LI Qiang, WANG Zhou. Reduced-Reference Image Quality Assessment Using Divisive Normalization-Based Image Representation[J]. IEEE Journal of Selected Topics in Signal Processing, 2009, 3(2): 201-211.

[41] WANG Zhou, WU Guixing, SHERIKH H.R, *et al*. Quality-Aware Images[J]. IEEE Transactions on Image Processing, 2006, 15(6): 1680-1689.

[42] MA Lin, LI Songnan, ZHANG Fan, *et al*. Reduced- Reference Image Quality Assessment Using Reorganized DCT-Based Image Representation[J]. IEEE Transactions on Multimedia, 2011, 13(4): 824-829.

[43] MA Lin, LI Songnan, NGAN K N. Reduced-Reference Video Quality Assessment of Compressed Video Sequences[J]. IEEE Transactions on Circuit System Video Technology, 2012, 22(10): 1441-1456.

[44] MA Lin, LI Songnan, ZHANG Fan, *et al*. Reduced-Reference Image Quality Assessment in Reorganized DCT Domain[J]. IEEE Transactions on Multimedia, 2011, 13(4): 824-829.

[45] WANG Zhou, BOVIK A C. Reduced- and no-Reference Image Quality Assessment: the Natural Scene Statistics Model Approach[J]. IEEE Signal Processing Magazine, 2011, 28(6): 29-40.

[46] LIANG Luhong, WANG Shiqi, CHEN Jianhua, *et al*. No-Reference Perceptual Image Quality Metric Using Gradient Profiles for JPEG 2000 [J]. Signal Processing: Image Communication, 2010, 25(7): 502-516.

[47] BRANDAO T, QUELUZ M P. No-Reference Image Quality Assessment Based on DCT Domain Statistics[J]. Signal Processing, 2008, 88(4): 822-833.

[48] WANG Zhou, SHEIKH H R, BOVIK A C. No-Reference Perceptual Quality Assessment of JPEG Compressed Images[C]// Proceedings of 2002 International Conference on Image Processing (ICIP 2002): September 22-25, 2002. Rochester, NY, USA, 2002: I-477-I-480.

[49] FERZLI R, KARAM L J. A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB)[J]. IEEE Transactions on Image Processing, 2009, 18(4): 717-728.

[50] SHEIKH H R, BOVIK A C, CORMARK L. No-Reference Quality Assessment Using Nature Scene Statistics: JPEG 2000[J]. IEEE Transactions on Image Processing, 2005, 14(11): 1918-1927.

[51] MITTAL A, MURALIDHAR G S, GHOSH J, *et al*. Blind Image Quality Assessment without Human Training Using Latent Quality Factors[J]. IEEE Signal Processing Letters, 2012, 19(2): 75-78.

[52] SAAD M A, BOVIK A C, CHARRIER C. A Perceptual DCT Statistics Based Blind Image Quality Metric[J]. IEEE Signal Processing Letters, 2010, 17(6): 583-586.

[53] MOORTHY A K, BOVIK A C. Blind Image Quality Assessment: from Natural Scene Statistics to Perceptual Quality[J]. IEEE Transactions on Image Processing, 2011, 20(12): 3350-3364.

[54] HEMAMI S S, REIBMAN A R. No-Reference Image and Video Quality Estimation: Applications and Human-Motivated Design[J]. Image Communication, 2010, 25(7): 469-481.

[55] Video Coding Experts Group (VCEG). Joint call for proposals on scalable video coding extensions of high efficiency video coding (HEVC) [R]. 2012.

[56] LEE J S, SIMONE F D, RAMZAN N, *et al*. Subjective Evaluation of Scalable Video Coding for Content Distribution[C]// Proceedings of the International Conference on Multimedia (MM'10): October 25-29, 2010. Firenze, France, 2010: 65-72.

[57] ZHAI Guangtao, CAI Jianfei, LIN Weisi, *et al*. Cross-Dimensional Perceptual Quality Assessment for Low Bit-Rate Videos[J]. IEEE Transactions on Multimedia, 2008, 10(7): 1316-1324.

[58] 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.

[59] YEGANEH H, WANG Zhou. Objective Quality Assessment of Tone Mapped Images[J]. IEEE

Transactions on Image Processing, 2013, 22(2): 657-667.
[60] YEGANEH H, WANG Zhou. Objective Quality Assessment of Tone Mapping Algorithms[C]// Proceedings of the 17th IEEE International Conference on Image Processing (ICIP 2010): September 26-29, 2010. Hong Kong, China, 2010: 2477-2480.
[61] AYDIN T O, MANTIUK R, MYSZKOWSKI K, *et al*. Dynamic Range Independent Image Quality Assessment[J]. ACM Transactions on Graphics, 2008, 27(3): 69.
[62] CADIK M, HAJDOK O, LEJSEK A, *et al*. Evaluation of Tone Mapping Operators[EB/OL]. http://www.cgg.cvut.cz/members/cadikm/tmo.
[63] LIU Jianqing, YANG Y. Multiresolution Color Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(7): 689-700.
[64] BORSOTTI M, CAMPADELLI P, SCHETTINI P C. Quantitative Evaluation of Color Image Segmentation Results[J]. Pattern Recognition Letter, 1998, 19(8): 741-747.
[65] ZHANG Hui, FRITTS J E, GOLDMAN S A. An Entropy-Based Objective Segmentation Evaluation Method for Image Segmentation[J]. SPIE - Electronic Imaging Storage and Retrieval Methods and Applications for Multimedia, 2004, 5307: 38-49.
[66] REN Xiaofeng, MALIK J. Learning a Classification Model for Segmentation[C]// Proceedings of the 9th IEEE International Conference on Computer Vision, 2003 (ICCV): October 13-16, 2003. Nice, France, 2003: 10-17.
[67] CHRISTENSEN H I, PHILLIPS P J. Empirical Evaluation Methods in Computer Vision[M]. Singapore: World Scientific Publishing Company, 2002.
[68] MARTIN D, FOWLKES C, TAL D, *et al*. A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics[C]// Proceedings of the 8th IEEE International Conference on Computer Vision, 2001 (ICCV 2001): July 7-14, 2001. Vancouver, BC, Canada, 2001: 416-423.
[69] FREXENET J, MUNOZ X, RABA D, *et al*. Yet Another Survey on Image Segmentation: Region and Boundary Information Integration [C]// Proceedings of the 7th European Conference on Computer Vision-Part III (ECCV'02): May 28-31, 2002. Copenhagen, Denmark, 2002: 408-422.
[70] MARTIN D R. An Empirical Approach to Grouping and Segmentation[D]. San Francisco, USA: University of California, Berkeley, 2002.
[71] RAND W M. Objective Criteria for the Evaluation of Clustering Methods[J]. Journal of the American Statistical Association, 1971, 66(336): 846-850.
[72] FOWLKES E B, MALLOWS C L. A Method for Comparing Two Hierarchical Clusterings[J]. Journal of the American Statistical Association, 1983, 78(833): 553-569.
[73] UNNIKRISHNAN R, HEBERT M. Measures of Similarity[J]. IEEE Workshop on Applications of Computer Vision, 2005, 1: 394-400.
[74] PENG Bo, ZHANG Lei. Evaluation of Image Segmentation Quality by Adaptive Ground Truth Composition[J]. Lectures Notes in Computer Science, 2012, 7574: 287-300.
[75] MARTIN D R, FOWLKES C C, MALIK J. Learning to Detect Natural Image Boundaries Using Local Brightness, Color and Texture Cues[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(5): 530-549.
[76] SAMPAT M P, WANG Zhou, GUPTA S, *et al*. Complex Wavelet Structural Similarity: A New Image Similarity Index[J]. IEEE Transactions on Image Processing, 2009, 18(11): 2385-2401.
[77] MOORTHY A K, SU Chechun, MITTAL A, *et al*. Subjective Evaluation of Stereoscopic Image Quality[J]. Signal Processing: Image Communication, 2012.
[78] LEBRETON P, RAAKE A, BARKOWSKY M, *et al*. Evaluating Depth Perception of 3D Stereoscopic Videos[J]. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(6): 710-720.
[79] BOSC E, PEPION R, LE CALLET P, *et al*. Towards a New Quality Metric for 3-D Synthesized View Assessment[J]. IEEE Journal of Selected Topics in Signal Processing, 2011, 5(7): 1332-1343.
[80] HEWAGE C T E R, MARTINI M G. Reduced-Reference Quality Assessment for 3D Video Compression and Transmission[J]. IEEE Transactions on Consumer Electronics, 2011, 57(3): 1185-1193.
[81] MOORTHY A K, BOVIK A C. A Two-Step Framework for Constructing blind Image Quality Indices[J]. IEEE Signal Processing Letters, 2011, 17(2): 587-599.
[82] BENOIT A, LE CALLET P, CAMPISI P, *et al*. Quality Assessment of Stereoscopic Images[J]. EURASIP Journal on Image and Video Processing, 2009: 1-13.
[83] HEWAGE C T E R, MARTINI M G. Reduced-Reference Quality Metric for 3D Depth Map Transmission[C]// Proceedings of 3DTV-Conference: The True Vision- Capture, Transmission and Display of 3D Video (3DTV-CON),

2010: June 7-9, 2010. Tampere, Finland, 2010: 1-4.
[84] YOU Junyong, XING Liyuan, PERKIS A, *et al*. Perceptual Quality Assessment for Stereoscopic Images Based on 2D Image Quality Metrics and Disparity Analysis[C]// Proceedings of the 5th International Workshop on Video Processing and Quality Metrics (VPQM 2010): January 13-15, 2010. Scottsdale, Arizona, USA, 2010: 281-316.
[85] GORLEY P, HOLLIMAN N. Stereoscopic Image Quality Metrics and Compression[J]. Stereoscopic Displays and Applications, 2008, 6803.
[86] SHEN Lili, YANG Jiachen, ZHANG Zhouyun. Stereo Picture Quality Estimation Based on a Multiple Channel HVS Model[C]// Proceedings of the 2nd International Congress on Image and Signal Processing, 2009 (CISP'09): October 17-19, 2009. Tianjin, China, 2009: 1-4.
[87] YANG Jiachen, HOU Chunping, ZHOU Yuan, *et al*. Objective Quality Assessment Method of Stereo Images[C]// Proceedings of 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009: May 4-6, 2009. Potsdam, Germany, 2009: 1-4.
[88] ZHU Zhongjie, WANG Yuer. Perceptual Distortion Metric for Stereo Video Quality Evaluation[J]. WSEAS Transactions on Signal Processing, 2009, 5(7): 241-250.
[89] AKHTER R, SAZZAD Z M P, HORITA Y, *et al*. No Reference Stereoscopic Image Quality Assessment[J]. Stereoscopic Displays and Applications, 2010, 7524: 752407-752407-12.
[90] RUBINSTEIN M, GUTIERREZ D, SORKINE O, *et al*. A Comparative Study of Image Retargeting[C]// Proceedings of the 3rd ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia: December 15-18, 2010. Hong Kong, China, 2010: 160.
[91] MA Lin, LIN Weisi, DENG Chenwei, *et al*. Image Retargeting Quality Assessment: A Study of Subjective Scores and Objective Metrics[J]. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(6): 626-639.
[92] MA Lin, LIN Weisi, DENG Chenwei, *et al*. Study of subjective and Objective Quality Assessment of Retargeted Images[C]// Proceedings of 2002 International Conference Symposium on Circuits and Systems (ISCAS 2012): May 20-23, 2012. Seoul, Korea, 2012: 626-639.
[93] MANJUNATH B S, OHM J, VASUDEVAN V V, *et al*. Color and Texture Descriptors[J]. IEEE Transactions on Circuits Systems for Video Technology, 2001, 11(6): 703-715.
[94] LIU Ce, YUEN J, TORRALBA A, *et al*. SIFT Flow: Dense Correspondence across Different Scenes[C]// Proceedings of the 10th European Conference on Computer Vision: Part III (ECCV'08): October 12-18, 2008. Marseille, France, 2008: 28-42.
[95] SIMAKOV D, CASPI Y, SHECHTMAN E, *et al*. Summarizing Visual Data Using Bidirectional Similarity[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2008 (CVPR 2008): June 23-28, 2008. Anchorage, Alsaka, USA, 2008: 1-8.
[96] BARNES C, SHECHTMAN E, FINKELSTEIN A, *et al*. PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing[C]// Proceedings of the International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH): August 3-7, 2009. New Orleans, Louisiana, USA, 2009: 24.
[97] PELE O, WERMAN M. Fast and Robust Earth Mover's Distances[C]// Proceedings of 2009 IEEE 12th International Conference on Computer Vision: September 29-October 2, 2009. Kyoto, Japan, 2009: 460-467.
[98] RUBNER Y, TOMASI C, GUIBAS L J. The Earth Mover's Distance as a Metric for Image Retrieval[J]. International Journal of Computer Vision, 2000, 40(2): 99-121.

Biographies

MA Lin, received his B.E. and M.E. degrees in computer science from Harbin Institute of Technology (HIT), China in 2006 and 2008, respectively. He is currently pursuing his Ph.D. degree at the Department of Electronic Engineering, Chinese University of Hong Kong (CUHK), Hong Kong, China. He was a recipient of the Best Paper Award in Pacific-Rim Conference on Multimedia in 2008. He was awarded Microsoft Research Asia Fellowship in 2011. His research focuses on image/video quality assessment and compression.

DENG Chenwei, received his Ph.D. degree in signal and information processing from Beijing Institute of Technology (BIT), Beijing, China in 2009. From May 2009 to August 2012, he was a Postdoctoral Research Fellow at the School of Computer Engineering, Nanyang Technological University (NTU), Singapore. Currently, he is an Associate Professor with the School of Information and Electronics, Beijing Institute of Technology, China. His research interests include image/video coding, perceptual modeling, and

computer vision. *The corresponding author. Email: cwdeng@bit.edu.cn

King Ngi Ngan, is a Chair Professor at the Department of Electronic Engineering, Chinese University of Hong Kong (CUHK), Hong Kong, China. He holds honorary and visiting professorships of numerous universities in China, Australia and South East Asia. His research interests are in video coding and visual signal processing, and he published extensively in these areas. Prof. Ngan is a Fellow of IEEE (USA), Institute of Educational Technology (UK), and Institution of Engineers Australia, and an IEEE Distinguished Lecturer in 2006-2007.

LIN Weisi, received his B.S. and M.S. degrees from Zhongshan University, China, and a Ph.D. degree in computer vision from King's College, London University, UK. He is now an Associate Professor in the School of Computer Engineering, Nanyang Technological University (NTU), Singapore. Prior to this, he was the Laboratory Head and the Acting Department Manager in the Institute for Infocomm Research, Singapore. His areas of expertise include perceptual modeling, video compression, multimedia quality evaluation and computer vision.