Overview of Quality Assessment for Visual Signals and Newly Emerged Trends

Lin Ma[†], Chenwei Deng[‡], King N. Ngan[†], and Weisi Lin[#]
[†]Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong
[‡]School of Information and Electronics, Beijing Institute of Technology, China
[#]School of Computer Engineering, Nanyang Technological University, Singapore
{lma,knngan}@ee.cuhk.edu.hk; cwdeng@bit.edu.cn; wslin@ntu.edu.sg

***Abstract*—Quality assessment is not only essential on its own for testing, optimizing, benchmarking, monitoring and inspecting related systems and services, but 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 the paper, we provide an overview of the quality assessment approaches for traditional visual signals, as well as the newly emerged ones, which covers the subjective quality evaluation and objective quality metrics of scalable and mobile videos, high dynamic range (HDR) images, image segmentation results, 3D images/videos, and retargeted images. Also the challenges for designing effective quality metrics and corresponding applications are discussed.**

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 and how the quality of visual signals is gauged dictates 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 is best judged by human subjects. However, subjective assessment 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 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, the video databases: LIVE, EPFL, and VQEG FR-TV phase I. Detailed information about these databases can be referred to [1]-[3].

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 [1]-[2], perceptual visual quality metrics are surveyed, which mostly focus on the quality assessment of traditional visual signals. However, recently more and more visual signals are 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 metrics 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 objective 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. In Section III, quality assessment of newly emerged visual signals will be reviewed, as well as the corresponding applications. Section IV will discuss the challenges and future trends for designing objective quality metrics. Finally, Section V will conclude the paper.

II. OBJECTIVE QUALITY ASSESSMENT FOR TRADITIONAL VISUAL SIGNALS

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. In the following, we are to review some representative methods in each category.

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong SAR, China (Project CUHK 415712), and in part by the Singapore Ministry of Education Academic Research Fund (AcRF) Tier 2 under Grant T208B1218.

A. FR Quality Metric

Mean square error (MSE), signal to noise ratio (SNR), and peak SNR (PSNR) are the simplest FR metrics, where the signal fidelity is measured. Although being widely accepted, these methods cannot capture the perceived quality accurately, especially when they are used in the case of non-additive distortion. In the past years, perceptual visual quality metrics (PVQMs) have been extensively investigated [4]-[17]. PVQM refers to the objective computational model for predicting human perceived visual quality, based upon the properties of human visual system (HVS). For images, the well cited SSIM [4] and its relatives [5] [6] exploit the structural sensitivity of human eyes, while human brain theory is utilized in [7] [8]. Another image metric is VIF [9], which attempts to relate image quality to the amount of information that is shared by the reference and test images. For videos, the well-known perception-based metrics include VDP [10], PDM [11], PEVQ [12], etc. In addition, the widely used PNSR has been modified by incorporating some properties of the HVS in order to improve its prediction performance. In [13] [14], a JND threshold is estimated, and a modified PSNR is computed by subtracting the JND threshold from the difference between the reference and distorted videos. More recently, MOVIE [15] integrates both spatial and temporal aspects of distortion assessment, and uses optical flow estimation to adaptively guide spatial-temporal filtering using 3-D Gabor filter banks. It should be noted that another class of metrics based on learning framework have also been proposed [17]. Basically, these models involve two stages (i.e., feature extraction, and feature pooling). In the first stage, some image/video features are extracted, and then machine learning is applied to obtain a trained metric.

B. RR Quality Metric

RR quality metrics are roughly classified into 3 different but related categories. The first approach is based on *modeling image distortions*. The quality metrics [19]-[20] 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 [21]-[23] are developed by considering the HVS properties. For example, in [22], 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 ANSI. The third type of approaches is based on *modeling natural image/video statistics*. The underline essential of these metrics [18] [24]-[27] 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 [24], generalized Gaussian density (GGD) is employed to depict the wavelet coefficient distribution. In [25] [27], 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.

C. NR Quality Metric

The design of NR quality metrics is an extremely difficult task. Therefore, most of NR metrics focused on the image/video degraded by the specific distortions. In [28], the image/video statistics are discussed for designing NR quality metrics. Nowadays, generic NR metric for images are researched [30]-[32], where machine learning is employed to fuse the quality metric designed for specific distortions together. In [29], a comprehensive overview on the NR quality metric has been provided, which can be referred to for detailed information. In this paper, due to the page limitation, we will not discuss this point in more details.

In this paper, we employ LIVE image/video database to evaluate the performances of the corresponding quality metrics, which are illustrated in Table I and Table II, respectively. Three statistical parameters, specifically linear coefficient correlation (LCC), Spearman-rank order correlation coefficient (SROCC), and root mean square prediction error (RMSE) are employed to evaluate the metric's performance. It can be observed that FR FSIM [5] and MOVIE [15] performs the best for image and video quality assessment.

Table I. Comparison of different image quality metrics

	Type	LCC	SROCC	RMSE
PSNR	FR	0.872	0.876	13.37
SSIM [4]	FR	0.904	0.910	11.68
VIF [9]	FR	0.960	0.963	7.673
FSIM [5]	FR	0.960	0.963	7.678
WNISM [24]	RR	0.738	0.779	18.43
RR Metric [25]	RR	0.883	0.879	12.84
NR Metric [30]	NR	0.823	0.800	-
BLIINDS [31]	NR	0.680	0.663	20.01
DIIVINE [32]	NR	0.917	0.916	10.90

Table II. Comparison of different video quality metrics

	Type	LCC	SROCC	RMSE
PSNR	FR	0.5398	0.5234	9.241
SSIM [4]	FR	0.4999	0.5247	9.507
MOVIE [15]	FR	0.8116	0.7890	-
VQM [22]	RR	0.7160	0.7029	7.664

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 will 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.

A. Scalable and mobile video quality assessment

In [33], the authors have performed a large scale study on the subjective quality of the scalable video coding (SVC) for content distribution. The influence of the combination of scalable parameters in SVC was studied, where the subjective study is performed in a pair comparison way. 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 built database has been made publicly available [33] for researchers working on objective quality design to measure the video sequences generated by SVC.

In [34], 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 impact of different display modes (mobile phones and tablets) was studied by comparing the corresponding subjective ratings of the mobile videos. Moreover, several representative quality metrics were evaluated on the mobile database to provide the benchmark results, which are illustrated in Table III.

Table III. Quality metric comparisons on the mobile database

	Tablet			Mobile		
	SROCC	LCC	RMSE	SROCC	LCC	RMSE
PSNR	0.5886	0.6348	0.6630	0.6780	0.6909	0.6670
SSIM [4]	0.4300	0.4893	0.7483	0.6498	0.6637	0.6901
VIF [9]	0.7261	0.7635	0.5541	0.7439	0.7870	0.5692
VQM [22]	0.5552	0.6150	0.6980	0.6945	0.7023	0.6663
MOVIE [15]	0.6792	0.7828	0.5342	0.6420	0.7157	0.6444

B. HDR images quality assessment

HDR images are becoming more widely available, due to recent advances in imaging and computer graphics technologies. Tone mapping operators 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 [35]-[37]. In [35] [36], the authors proposed an objective quality metric by combining the multi-scale signal fidelity measure based on a modified SSIM and a naturalness measure based on natural image intensity statistics. With validation on the subject-rated image database [46], 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 [37], 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. Moreover, the metric can be applied for the evaluation of direct and inverse tone mapping operators.

Table IV. Performance of HDR quality metric [35] [36]

	Structural fidelity	Naturalness	Overall
KROCC	0.6923	0.3846	0.7179
SROCC	0.7912	0.5385	0.8187

C. Image segmentation quality assessment

In [38], 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. With evaluation on the benchmark Berkeley segmentation database, the method can faithfully reflect the perceptual quality of the segmentation.

D. 3D images and videos

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). Therefore, many researchers have devoted their efforts for creating benchmark database and quality metrics of 3D visual signals [39]-[42]. In [39], the authors created a 3D image database that incorporates ‘true’ depth information along with stereoscopic pairs and human opinion scores. Also a variety of 2D and 3D quality models are evaluated based on the created database. It is claimed that 3D quality assessment should focus on key unresolved problems including stereoscopic distortions, 3D masking, and so on. As we all know, the depth information is very important for the 3D video application, such as displaying, rendering, and so on. In [40], 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 [41] 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 [42] 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.

Table V. Metric performances on image retargeting database

	EH	E MD	BSE	SIFT-flow
LCC	0.3422	0.2760	0.2896	0.3141
SROCC	0.3288	0.2904	0.2887	0.2899
RMSE	12.686	12.977	12.922	12.817

E. 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 public image retargeting quality databases [43]-[45]. In [43], the authors compared the retargeted image quality to figure out which retargeting method performs the best. In [44] [45], the authors present the results of a subjective study on the perceived quality of image retargeting and make the resultant database public. The MOS value for each retargeted image is analyzed with three aspects, namely, retargeting ratio, retargeting algorithm and visual content. Several existing quality metrics on retargeted images, such as edge histogram (EH), SIFT-flow, bidirectional similarity (BDS), and earth’s mover distance (EMD), have been evaluated on the database as shown in Table V.

IV. DISCUSSION

Although we have discussed the quality metrics for traditional and newly emerged visual signals, there are still many challenges remaining in designing accurate objective quality metrics. Firstly, visual signals have various contents leading to different distortion masking levels. Secondly, the visual signals before being perceived will pass through numerous processing stages, e.g., compression, transmission, etc. Each step will introduce different types of artifacts. Thirdly, the visual signals will be interpreted by the HVS system, the physiological knowledge of which is limited. Fourthly, quality assessment is viewer dependent, related to unpredictable factors like the viewer's interests, expectations, quality experience, etc. Therefore, it is impossible for an objective quality metric to match individual subjective ratings unconditionally.

For designing any scientific visual signal processing systems, a certain quality criterion or metric should be involved explicitly or implicitly. If the quality criterion is available, it may be used to not only assess the performance of these systems, but also optimize them to obtain the best results under this criterion. However, current quality metric design tends to be infeasible for direct incorporation into practical applications. The reason is that most of these metrics include many components, such as HVS modeling, signal statistical property, machine learning, and so on. Therefore, they are not presented in a simple and convex form, leading to difficulties in optimization. In this manner, they cannot be easily to be incorporated into the multimedia processing systems.

Furthermore, the applications and quality metrics are believed to mutually benefit each other. Quality metrics can definitely benefit the applications. On the other hand, new challenges arising from real applications, such as the newly emerged visual signals (e.g. image retargeting, 3D images/vides, HDR images, and so on), will impact the new development of future quality metrics.

V. CONCLUSION

In this paper, we provide an overview of objective quality metrics for visual signals, specifically the traditional and newly emerged visual signals, consisting of the scalable and mobile video, HDR image, image segmentation, 3D images/video, and image retargeting quality assessment. Moreover, the challenges for designing quality metrics and developing corresponding applications are discussed. It is believed that the metrics and applications will mutually benefit each other.

REFERENCES

[1] W. Lin, and C.-C. Jay Kuo, "Perceptual visual quality metrics: A suvey", *J. Vis. Commun. Image R.*, 22(2011), pp: 297-312.

[2] T.-J. Liu, W. Lin, and C.-C. Jay Kuo, "Recent developments and future trends in visual quality assessment", *APSIPA ASC* 2011.

[3] S. Winkler, "Analysis of public image and video databases for qulaity assessment", *J-STSP*, vol. 6, no. 6, pp. 616-625, Oct. 2012.

[4] Z. Wang, et al., "Image qualityassessment: from error visibility to structural similarity", *TIP*, vol. 13, no. 4, pp. 600–612, 2004.

[5] L. Ma, S. Li, and K. N. Ngan, "Visual horizontal effect for image quality assessment", *IEEE Signal Process. Letters*, vol. 17, no. 7, pp. 627-630, Jul. 2010.

[6] A. Liu, W. Lin, M. Narwaria, "Image quality assessment based on gradient similarity", *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500 - 1512, 2012

[7] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments", *TMM*. vol. 13, pp. 935-949, 2011.

[8] J. Wu, W. Lin, G. Shi, and A. Liu, "A perceptual quality metric with internal generative mechanism", *IEEE Trans. Image Process.*, accepted.

[9] H. R. Sheikh, and A. C. Bovik, "Image information and visualquality, *IEEE Trans. Image Processing*, vol. 15, no. 2, pp.430–444, Feb. 2006.

[10] S. Daly, The visible difference predictor: an algorithm for the assessment of image fidelity, Digital Images and Human Vision, MIT Press, 1993, pp. 179−206.

[11] S. Winkler, in: Digital Video Quality: Vision Models and Metrics, John Wiley & Sons, 2005.

[12] ITU-T Recommendation J.247, Objective perceptual multimedia video quality measurement in the presence of a full reference, Aug. 2008.

[13] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Just noticeable distortion model and its applications in video coding", *SPIC.,* vol. 20, pp. 662−680, 2005.

[14] L. Ma, K. N. Ngan, F. Zhang, and S. Li, "Adaptive Block-Size Transform Based Just-Noticeable Difference Model for Images/Videos", *SPIC*, vol. 26, pp. 162-174, 2011.

[15] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos", *TIP*, vol. 19, no. 2, Feb 2010.

[16] F. Zhang, L. Ma, S. Li, and K. N. Ngan, "Practical Image Quality Metric Applied to Image Coding", TMM, vol. 13, no. 4, pp. 615-624, Aug. 2011.

[17] M. Narwaria, and W. Lin, "SVD-based quality metric for image and video using machine learning", *TSMC-- B*,Vol. 42(2), pp. 347 - 364, 2012.

[18] Q. Li, et al., "Reduced-reference image quality assessment using divisive normalization-based image representation", *J-STSP*, pp. 201-211, Apr. 2009.

[19] S. Wolf, and M. H. Pinson, "Low bandwidth reduced reference video quality monitoring system", *QoMEX*, Jan. 2005.

[20] M. Tagliasacchi, *et. al*, "A reduced-reference structural similarity approximation for videos corrupted by channel errors", *Multimedia Tools Appl.*, vol. 48, no. 3, pp. 471-492, Jul. 2010.

[21] P. Le Callet, C. V. Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment", *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1316-1327, May. 2006.

[22] M. H. Pinson, and S. Wolf, "A new standardized method for objectively measuring video quality", *IEEE Trans. Broadcasting*, vol. 50, pp. 312-322, Sept. 2004.

[23] J. A. Redi, et al., "Color distribution information for reduced-reference assessment of perceived image quality", *TCSVT*, vol. 20, pp. 1757-1769, Dec. 2010.

[24] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. Yang, and A. C. Bovik, "Quality-aware images", *TIP*, vol. 15, no. 6, pp. 1680-1689, Jun. 2006.

[25] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation", *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824-829, Aug. 2011.

[26] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference video quality assessment of compressed video sequences", *TCSVT,* vol. 22, no. 10, pp. 1441-1456, Oct. 2012.

[27] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference image quality assessment in reorganized DCT domain", *Signal Processing: Image Communication*, accepted.

[28] Z. Wang et al., "Reduced- and no-reference image quality assessment: the natural scene statistic model approach," *SPM,* vol. 28, no. 6, pp. 29-40, Nov. 2011.

[29] S. S. Hemami, et al., "No-reference image and video quality estimation: applications and human-motivated design", *SPIC*, vol. 25, no. 7, pp. 469-481, Aug. 2010.

[30] A. Mittal, et al. "Blind image quality assessment without human training using latent quality factors", *SPL*, vol. 19, no. 2, pp. 75-78, Feb. 2012.

[31] M. A. Saad, A. C. Bovik, and C. Charrier, "A perceptual DCT statistics based blind image quality metric", *SPL*, vol. 17, pp. 583-586, 2010.

[32] A. K. Moorthy, and A.C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality", *TIP*, pp. 3350-3364, 2011.

[33] J. S. Lee, *et al.*, "Subjective evaluation of scalable video coding for content distribution", *ACM MM*, 2010.

[34] A. K. Moorthy, et al., "Video quality assessment on mobile devices: subjective, behavioral and objective studies", *J-STSP*, vol. 6, no. 6, pp. 652-671, Oct. 2012.

[35] H. Yeganeh, and Z. Wang, "Objective quality assessment of tone mapped images", *TIP*, accepted.

[36] H. Yeganeh, and Z. Wang, "Objective quality assessment of tone mapping algorithms", ICIP, 2010.

[37] T. O. Aydin, et al., "Dynamic range independent image quality assessment", *SIGGRAPH*, 2008.

[38] B. Peng, and L. Zhang, "Evaluation of image segmentation quality by adaptive ground truth composition", *ECCV*, 2012.

[39] A. K. Moorthy, C. C. Su, A. Mittal and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *SPIC*, 2012, accepted.

[40] P. lebreton, *et al.*, "Evaluating depth perception of 3D stereoscopic videos", J-STSP, vol. 6, no. 6, pp. 710-720, Oct. 2012.

[41] E. Bosc, *et al.*, "Towards a new quality metric for 3-D synthesized view assessment", J-STSP, vol. 5, no. 7, pp. 1332-1343, Nov. 2011.

[42] C. T. E. R. Hewage, et al., "Reduced-reference quality assessment for 3D video compression and transmission", *TCE,* vol. 57, no. 3, pp. 1185-1193, Aug. 2011.

[43] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," in Proc. *SIGGRAPH Asia*, 2010.

[44] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: a study of subjective scores and objective metrics", *J-STSP,* vol. 6, Oct. 2012.

[45] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Study of subjective and objective quality assessment of retargeted images", *ISCAS* 2012.

[46] M. Cadik et al., "Evaluation of tone mapping operators," http://www.cgg.cvut.cz/members/cadikm/tmo.