A Survey of Image Quality Measures

Kim-Han Thung
Dept. of Electrical Engineering
University of Malaya
Lembah Pantai, 50603 Kuala Lumpur

Paramesran Raveendran
Dept. of Electrical Engineering
University of Malaya
Lembah Pantai, 50603 Kuala Lumpur

***Abstract*—Image quality assessment is one of the challenging field of digital image processing system. It can be done subjectively or objectively. PSNR is the most popular and widely used objective image quality metric but it is not correlate well with the subjective assessment. Thus, there are a lot of objective image quality metrics (IQM) developed in the past few decades to replace PSNR. This paper provides a literature review of the current subjective and objective image quality measures. The purpose of this paper is to collect reported quality metrics and group them according to their strategies and techniques.**

I. Introduction

With the help of advanced technology in micro processor which provide faster computation with cheaper cost, digital imaging system has proliferated in the past few decades and become the most common form of image processing in 2000s. In a typical digital imaging system, the image is captured and transformed into digital signal by the sensor (in the camera). This raw digital image signal is then processed to reduce noise and compressed for storage or transmission. When the image is finally displayed on the screen to the end user, it might not be as same as the original version because it has been exposed to various kind of distortions. The sources of distortions could be ranged from motion blurring, gaussian noise, sensor inadequacy, compression, error during transmission and so on. It is very important for an digital imaging system to be able to predict the quality of the image, so that it can maintain, control and possibly enhance the quality of the image before storage or transmission. Hence, a realiable image quality metric (IQM) is very crucial in the development of image and video signal processing systems.

There are basically two types of image quality assessment (IQA) techniques, namely the subjective method, which involve human beings to evaluate the quality of the images, and the objective method, which compute the image quality automatically (please refer to Fig. 2). Since human beings are the ultimate users of most of the multimedia applications, subjective evaluation is perhaps the most accurate and reliable way of assessing the quality of an image. However, this method is too slow, inconvenient and expensive for practical usage. Thus, objective image quality metrics that can automatically predict the perceived image quality come in handy. The goal of the objective IQA research is thus to predict the quality of an image as closely as to the subjective assessment.

In the past few decades, a lot of works have been done on IQA. The rapid growth in the research of IQA is motivated by the area of applications below [1]:

1) IQM can be used to *benchmark* image and video processing system and algorithm, such as compression, restoration, denoising and deblurring techniques. For example, it's always the goal of the compression algorithm to achieve higher compression ratio at the desire level of quality. A reliable IQM can be used to determine which algorithm provide the best result.
2) IQM can be used to *monitor* image quality in the quality control systems. For example, in the video streaming application, IQM can be used to examine the quality of the video signal in transmission, so that the encoder can adjust its parameter accordingly to transmit optimum quality video signal using least amaunt of bandwidth.
3) IQM can be embedded into image and video processing system to *optimize* the algorithm and parameter settings. For example, in the visual communication system, it can help in setting optimal compression parameter at the encoder and optimize the reconstruction and denoising algorithm at the decoder.

In this paper, we collect various types of image quality metrics and group them according to their strategies and techniques. We focus on categorizing the objective IQM while briefly report the subjective metric.

The organization of this paper is given as follow. Section II describes the philosophy of image quality assessment. Section III describes the subjective IQA method, whereas section IV describes and summarizes the objective IQM. We conclude our paper in section VI.

II. Definition of Image Quality

Historically, image quality is described in terms of the visibility of the distortions in an image, such as colour shifts, blurriness, Gaussian noise and blockiness. The most common way of modeling an image quality metric is therefore by quantification of the visibility of these distortions. For example, Just Noticeable Difference (JND) model by Sarnoff [2] predicts subjective rating of an image by examining the visibility of distortions. In [3], Janssen has proposed a new philosophy for image quality. He regards image as carriers of visual information instead of two-dimensional signals, regards visual-cognitive processing as information processing rather than signal processing and regards image quality as image adequacy in the visual interaction process instead of visibility of distortions. Though this concept is interesting and seem

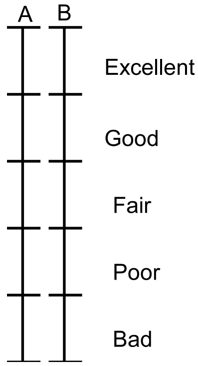

Fig. 1: Sample linear 5-segments grading scale used by DSCQS [4], [5]. The subject is asked to mark the impression of the image "A" and "B" on the scale. Either "A" or "B" is the source image while the other is the distorted version.

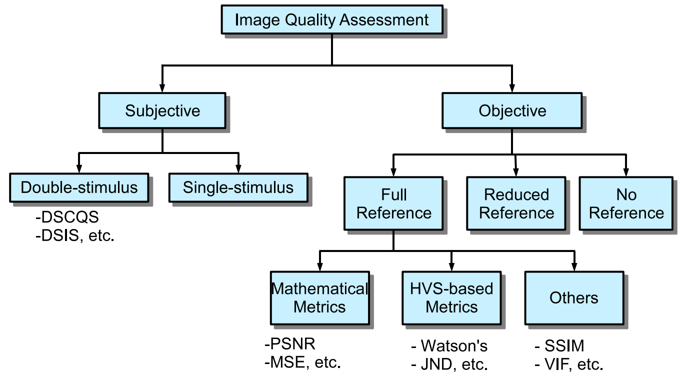

Fig. 2: Image Quality Assessment

applicable, most of the works are still reported based on the former concept, due to its simplicity and good performance.

III. Subjective Evaluation

Most of the current subjective experiments for image quality assessment are done according to *Rec. ITUR BT.500-10* [4]. In general, subjective evaluation of image quality could be *double stimulus* or *single stimulus* depending on the availability of the source (perfect quality) image. In double-stimulus methodology, subject is presented with the source and test images before evaluates their qualities on a linear quality scale as in Fig. 1. There are several double-stimulus methodologies [4] based on how the source image is presented to the subject, including the Double-Stimulus Continuous Quality Scale (DSCQS) and Double-Stimulus impairment scale (DSIS). In single-stimulus methodology, the subject evaluates the quality of the test images on a linear quality scale without the source as reference. The scores evaluated by multiple subjects are averaged for each test image to obtain mean opinion score (MOS) and difference mean opinion score (DMOS) [4], [5] after subject and outlier rejection for Single Stimulus and Double Stimulus methodology respectively. The subjective scores are used to evaluate the performance of the objective IQM in the next section. 5 publicly available subject-rated image databases, including LIVE database, Cornell A57 database, IVC database, Toyama database and TID2008 database, are listed in Table I together with their descriptions.

IV. Objective Metric

The objective IQA can be classified into full-reference (FR), reduced-reference (RF) and no-reference IQA based on the availability of the reference (ideal) image. In this paper, we are focusing on the FR objective image quality metric, where the quality of the distorted test images are obtained based on the comparison with the reference image which is assumed to be perfect in quality. Thus, the more appropriate term for the FR quality assessment is image "fidelity" measure. The image fidelity measures are grouped according to their strategies as in the following subsections.

A. Mathematical metrics

In this approach, the image is regarded as a 2D signal, and the dissimilarity (or similarity) between the reference and the distorted images is calculated as distortion (or quality) measure. Minkowski metric, for example, calculate the L_γ distance between the reference image $\mathbf{x}$ and distorted image $\mathbf{y}$ by:

$$E_\gamma = \left(\frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|^\gamma \right)^{1/\gamma}, \tag{1}$$

where x_i and y_i are the i-th samples in image $\mathbf{x}$ and $\mathbf{y}$ respectively, N is the number of image samples, and γ is in the range of $\gamma \in [1\ \infty)$. For $\gamma = 2$, one obtains the well-known Mean Square Error (MSE) formula if ignoring the square root:

$$E_2 = \left(\frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|^2 \right)^{1/2} = \sqrt{MSE}. \tag{2}$$

MSE and its variants such as peak signal-to-noise ratio (PSNR) are commonly used as objective IQM. PSNR is defined as:

$$PSNR = 10log_{10}\left(\frac{255^2}{MSE} \right) \tag{3}$$

where 255 is the maximum gray level of a 8bits/pixel monotonic image. Other mathematical distortion measures such as average difference, maximum difference, absolute error, Peak MSE, Laplacian MSE, etc can be found in [13]. In [13], [14], some correlation based measures that calculate the similarity between the reference and test images are described and compared. These metrics include structural content (SC), normalized cross-correlation (NCC), correlation quality, Czekanowski distance, etc. The major advantages of these metrics are its simplicity and mathematical tractability, but they are not correlating well with perceived quality measurement [15], [16] because the *Human Vision System* (HVS) characteristics are not considered in their models.

B. HVS based metrics

In this approach, the error signal (difference between the reference and the test images) is normalized according to its

TABLE I: Image databases with subjective evaluation

Database	Types of Distortions	Descriptions
LIVE database [6]	JPEG, JPEG2000, Gaussian Blur, White Noise and Fast Fading.	Total of 982 colour images (779 test images) were derived from 29 high quality colour images, 7 experiments are done by different number of subjects (22.8 on average). Single-stimulus methodoly was used in 7 experimental session and another experiment was done using double-stimulus methodology [4], [5] for realignment purpose.
IVC database [7]	JPEG, JPEG2000, LAR coding and Blurring.	Total of 235 distorted images were generated from 10 original colour images, 15 subjects were involved in the evaluation using DSIS method [4].
Toyama database [8]	JPEG and JPEG2000.	Total of 168 test images (84 for each distortions) were derived from 14 colour images, 16 subjects were involved in the evaluation using single-stimulus methodology [4].
A57 database [9]	JPEG, JPEG2000, JPEG2000 with DCQ algorithm, Gaussian Blur, White Noise and "Flat" allocation.	Total of 54 8-bitgrayscale test images were derived from 3 natural images and evaluated by 7 imaging expert subjects. In the experiment, subjects were instructed to arrange the distorted images such that the physical displacement between each distorted image and the original was linearly proportional to their subjective assessment of distortion [10].
TID2008 database [11]	17 types of distortions including Gaussian Blur, Gaussian Noise, JPEG, JPEG2000, Mean shift, transmission error, etc.	Total of 1700 distorted images were derived from 25 reference images. MOS were obtained from experiments done by 838 subjects from three country (Finland, Italy, and Ukraine). The methodology used was pair-wise sorting where the subjects were required to choose the best image that visually differed less from the original between two distorted images [12].

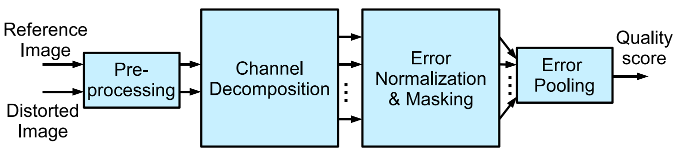

Fig. 3: A typical framework of the HVS based IQA system. The pre-processing stage before the channel decomposition process includes operations like alignment, color-space transformation, point-spread function (PSF) low-pass filtering (to simulate PSF of the eye optics), etc. [1].

visibility, as determined by psychophysics of human perception. Some of the HVS features that commonly used in IQM are:

1) contrast sensitivity function (CSF). Human perception is more sensitive to lower spatial frequency than higher one. Thus, some IQM models implement CSF as the low-pass (or band-pass) filter, and some implement CSF as weighting factors for subbands after frequency decomposition [1].
2) luminance contrast sensitivity. Human eyes are sensitive to luminance contrast rather than the absolute luminance value. According to Weber's law, the ratio $\Delta L/L$ of just noticeable luminance difference ΔL and the luminance L is constant for a wide range of luminances. For low background luminance (e.g. dark), this ratio is increased as the background luminance decreased. This effect is modeled as luminance masking in various IQM [17].
3) contrast masking. Contrast masking refers to the reduction in the visibility of one image component by the presence of another.

The general framework of the HVS based perceptual metric is shown in Fig. 3. The HVS features described above are implemented at the pre-processing and the error normalization and masking stage. The "channel decomposition" stage is used to transform image pixel value into independent (or at least decorrelated) spatial subbands, as this can improve the performance of quality metric [18]. Some of the transforms proposed are cortex transform, steerable pyramid transform, wavelet transform and DCT transform. The details of other blocks in the framework is given in [1].

The first attempt to use HVS in image fidelity measure is given by Mannos and Sakrison [19], and extended by many other researchers. Some of the representative models include "visible differences predictor" by Daly [20], model by Lubin [21], DCT-based and Wavelet-based Metric by Watson [17], "perceptual image distortion" by Teo and Heeger [22] and Just-Noticeable-Distortion (JND) model by Sarnoff [23]. A review of their metrics can be found in [24].

Though this approach is nearly universal accepted, there are some limitations, such as complexity and nonlinearity of HVS and suprathreshold problem [25].

C. Other Metrics

On the other hand, Wang *et al.* have proposed a new framework for the design of image quality measures in [25] and [26], the structural similarity approach. In structural similarity approach, it is assumed that HVS is highly adapted to the natural scenes information, which is highly structured. Thus, a measure of structural information change (between the reference and the distorted image) should provide a good approximation to perceived image distortion. They formulate the structural similarity as mean of SSIM index [18]:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4)$$

where $\{\mu_x, \sigma_x\}$ and $\{\mu_y, \sigma_y\}$ denote the mean intensity and standard deviation set of image block **x** and image block **y**, respectively, while σ_{xy} denote their cross correlation. C_1 and C_2 are small constants value to avoid instability problem when the denominator is too close to zero.

In [27], [28], the mutual information between the test and the reference images, is quantized to relate with visual quality, in which they called it visual fidelity information (VIF). A review of the recent works by Wang and Sheikh *et al.* can be found in [16].

V. Conclusion

This paper provides literature review of the current state-of-the-art image quality assessment. The methodology of subjective IQA and the image database that available to public are briefly overviewed. The objective quality metric discussed here are focused on the full-reference IQM, where the reference image is available to compute the distortion level of the test image. The definition of the image quality is discussed in this paper as well, as this can affect the design of the quality metric. A lot of mathematical metrics are defined by regarding IQA simply as 2D signal processing. Though of its simplicity, this method is not correlated well with the perceptual quality. By incorporating the HVS features into the metric, the performance of the IQM can be improved. A lot of HVS-based metrics have been proposed, trying to model the HVS into mathematical form, and normalize the error signal according to its visibility. HVS-based metric give a better correlation with the human perception but face some problems like complexity in design and difficulty in setting threshold for the visibility. Structural similarity on the other hand has model the IQA as the change in structural information in the image. It has low computation complexity and yet correlate well with subjective rating. There are some other metrics that have been proposed recently by defining image quality in other ways, such as in term of information carrier, communication theory, etc. Some of the future works that can be done on IQA including a faster and more accurate image fidelity prediction, no-reference quality assessment, modification of these algorithms for video quality assessment, etc.

References

[1] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *The Handbook of Video Databases: Design and Applications*, B. Furht and O. Marqure, Eds. Laboratory for Image and Video Engineering (LIVE), The University of Texas at Austin, Austin, TX 78712: CRC Press, Sept 2003, ch. 41, pp. 1041–1078.

[2] "Video clarity website," http://www.videoclarity.com/WhitePapers.html.

[3] R. Janssen, *Computational Image Quality*. SPIE Press, 2001.

[4] "Methodology for the subjective assessment of the quality of television pictures," Recommendation ITU-R BT.500-10, 1998.

[5] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii," http://www.vqeg.org/, Tech. Rep., 2003.

[6] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2." [Online]. Available: http://live.ece.utexas.edu/research/quality

[7] P. Le Callet and F. Autrusseau, "Subjective quality assessment irccyn/ivc database," 2005. [Online]. Available: http://www.irccyn.ec-nantes.fr/ivcdb/

[8] Z. M. P. Sazzad, Y. kawayoke, and Y. Horita, "Image quality evaluation database." [Online]. Available: http://mict.eng.u-toyama.ac.jp/database_toyama/

[9] D. Chandler and S. Hemami, "A57 image database," 2007. [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html

[10] D. M. Chandler and S. S. Hemami, "online supplement to 'vsnr: A visual signal-to-noise ratio for natural images based on near-threshold and suprathreshold vision'," 2007.

[11] N. N. Ponomarenko, "Tampere image database 2008, version 1.0." [Online]. Available: http://www.ponomarenko.info/tid2008.htm

[12] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, and F. Battisti, "Color image database for evaluation of image quality metrics," in *Proceedings of International Workshop on Multimedia Signal Processing*, Oct 2008, pp. 403–408.

[13] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on communications*, vol. 43, no. 12, pp. 2959–1965, 1995.

[14] Ismail Avcıbaş, "Image quality statistcs and their use in steganalysis and compression," Ph.D. dissertation, Boğaziçi University, 2001.

[15] B. Girod, "What's wrong with mean-squared error," in *Digital images and human vision*, A. Watson, Ed. Cambridge, MA, USA: MIT Press, 1993, ch. 15, pp. 207–220.

[16] Z. Wang and A. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[17] A. B. Watson, "Dct quantization metrices visually optimized for individual images," in *Human Vision, Visual Processing, and Digital Display IV*, ser. Proc. SPIE, B. E. Rogowitz, Ed., vol. 1913, 1993, pp. 202–216.

[18] Z. Wang, A. C. Bovik, and E. P. Simoncelli1, "Structural approaches to image quality assessment," in *Handbook of Image and Video Processing*, A. Bovik, Ed. Laboratory for Computational Vision, New York University, New York, NY 10003; Laboratory for Image and Video Engineering (LIVE), The University of Texas at Austin, Austin, TX 78712: Academic Press, May 2005, ch. 8.3.

[19] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, Jul 1974.

[20] S. Daly, *Digital Image and Human Vision*. MIT Press, 1993, ch. The visible difference predictor: an algorithm for the assessment of image fidelity, pp. 179–206.

[21] J. Lubin, *Digital Images and Human Vision*. MIT Press, 1993, ch. The use of psychophysical data and models in the analysis of display system performance, pp. 163–178.

[22] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *ICIP-94*, vol. 2. ICI, 1994, pp. 982–986.

[23] J. Lubin, *A visual discrimination model for imaging system design and evaluation*. World Science Publishing Co. Pte. Ltd., 1995, ch. 10, pp. 245–283.

[24] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in *in Handbook of Image and Video Processing*. Academic Press, 2000, pp. 669–684.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[26] Z. Wang and A. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81–84, Mar 2002.

[27] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2117–2128, 2005.

[28] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, 2006.