IMAGE QUALITY ASSESMENT-A REVIEW

Jagroop Singh[1]

[1]*Faculty DAV Institute of Engineering and Technology, Kabir Nagar, jalandhar.*

Abstract -*In most storage systems designed for images compression technologies are used to save memory. For instance, digital cameras use lossy compression technology for saving memory. In other words, a captured image will be compressed and then saved to memory card. The compression process will cause some details in the original image to disappear and thus causing artifacts. Consequently, if more compression is applied to an image, more space is saved but the compressed image has lower quality and more artifacts. Also in order to compare the results of the different post-processing algorithms image quality assessment is necessary.*

Methods for image quality evaluation can be generally classified as objective and subjective. By objective measures, some calculated parameters are calculated to indicate the reconstructed image fidelity and by subjective measure, viewers read images directly and make decisions based on their quality and intelligibility. In contrast to objective criteria, the subjective criterion is a complex and observer dependant evaluation method.

Key Words: **Compression, Post-Processing, Quality, Subjective**

1. INTRODUCTION

The obvious way of measuring image or video quality is to solicit human opinion. This is known as *subjective* quality assessment method and the average opinion about quality of a group of human subjects is sought. Basic fidelity measures including mean-squared error (MSE), peak signal-to- noise ratio (PSNR) are simple and widely used, but they do not always correlate well with perceived quality. Additionally, these measures require a reference that exists in the form of an "original" to compare with, which restricts their usability. Thus, reliable automatic methods for visual quality assessment are needed. Ideally, such a quality assessment system would "perceive" and measure image or video impairments just like a human being does.

2. SUBJECTIVE EVALUATION CRITERIA

In subjective testing, human observers are generally asked to rate image quality in terms of annoyance, where annoyance is a measure of how 'bad' the observer thinks the impairment is. The annoyance value correlates with the strength of the impairment as shown in Table 1.

Table 1: Five point scale of ITU Rec. 500-3 subjective assessment of images ITU-R Recommendation BT.500-3. (2002).

Score	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

There is wide variety of possible methods of subjective evaluation based on the ITU standards BT500 and P910 by ITU-R Recommendation BT.500-11. (2002), ITU-T Recommendation P.910. (1996). Subjective image quality metrics can generally be classified as:

- Mean opinion score (MOS)
- Single stimulus method
- Comparison method
- Double stimulus method
- Double stimulus continuous quality scale (DSCQS)

2.1. Mean Opinion Score (MOS)

The Mean Opinion Score (MOS) is a subjective error measure and is calculated by averaging the annoyance level for all observers. The perception based subjective evaluation, quantified by Mean opinion score was suggested by Grgic et al. (2004) [1]. For the set of distorted images, the MOS values were obtained from an experiment involving non-expert viewers. The original source image without compression was used as the reference condition. The assessor is asked to vote for the second keeping in mind the first. The method uses the five-grade impairment scale with proper description or each grade: 5-excellent quality, 4- good quality, 3- acceptable, 2- poor quality, 1- unacceptable quality Horita et al. (2006) [2]. At the end of the series of sessions, MOS for each test condition and test image are calculated as follows:

$$MOS(j) = \frac{1}{n}\sum_{i=1}^{n} S(i, j) \qquad (1)$$

where *n* denotes the number of observers and *S(i, j)* is the score given by the i^{th} observer to image *j.*

2.2 Single Stimulus Method

In the single stimulus method, the subject is presented with a single test object one at a time. At the end of each presentation, the subject is asked to give a rating for the test object. Then, the same procedure is repeated until all test objects are presented. In single stimulus method, the subject does not refer back to the previous assessment results for references as proposed by Tan et al. (1998) [3]. This method is typically used in experiments in which it is difficult to assess more than one stimulus at a time (e.g. audio and video assessments) or when the assessment time permitted is limited. The phenomenon of adaptation will tend to have a significant effect on the test results.

2.3 Comparison Method

In the comparison method, test objects are presented to the subject at the same time. During the presentation, the subject will have the opportunity to compare and sort the qualities of all test objects. The subject will rate the objects after they have been sorted. In this method, the effect of adaptation is least significant by Tan et al. (1998)[3].

2.4 Double Stimulus Method

Similar to the single stimulus method, the test objects of the double stimulus method are presented in a sequence as suggested by Narita (1994) [4]. But in each presentation, a constant reference object is also present at the same time. The subject is not informed about which is the reference object and is required to give rating for both the reference and test objects. This method is very popular in video assessment. It is relatively more time consuming, but the result yielded is less adaptive and more reliable than the single stimulus method.

2.5 Double Stimulus Continuous Quality Scale (DSCQS)

Double stimulus continuous quality scale (DSCQS) is a method in which source and processed image or video clips are presented in pairs to observers as suggested by Bovik (2003) [5]. The video or image presentation sequence is randomized. Viewers grade the quality of each clip then the data is processed in pairs. Until very recent times subjective measures such as MOS and DSCQS offered the most reliable quality measures.

Some issues that arise with subjective assessment include the cost and the fact that these methods cannot be used to monitor video quality in real time or continuously. The process requires special equipment and many people. Traditional analogue objective measurements, while still necessary, are not adequate to measure the quality of systems using digital compression. Thus the objective methods incorporating the characteristics of the human visual system including perceptual processes as proposed by Knee (2000) [6] should be used.

3. OBJECTIVE IMAGE QUALITY ASSESSMENT

Digital images suffer a wide variety of distortions in many image processing applications from compression to printing. Because of these the perceptual quality of the images are degraded. Therefore perceptual image quality measurement is important in many image processing applications. Through the subjective test is considered to be the most accurate method since it reflects human perception, it is time consuming and expensive. Furthermore, it cannot be done in real time. As a result objective image quality assessment methods are getting more attraction. O*bjective* quality assessment methods analyze images and videos and report their quality without human involvement. Such methods could eliminate the need for expensive subjective studies as suggested by Sheikh (2004)[7].

Objective assessment methods which serve as computational alternatives for expensive image quality assessment by human subjects, aimed at predicting perceived image quality aspects automatically and quantitatively. They are of fundamental importance to a broad range of image and video processing applications, such as for the optimization of image and video coding or for real time quality monitoring and control in displays.

In the last decades, a considerable research has been carried out on developing As suggested by Zhang (2006) [8]objective image quality metrics, can be generally classified into three categories: full- reference (FR) metrics, no-reference metrics (NR) and reduced-reference (RR) metrics.

3.1 Objective Full-Reference Quality Assessment (FRQA)

Researchers in the field of image quality assessment have attempted to measure quality using the so-called full-reference framework. This framework is a consequence of our limited understanding of human perceptions of quality. It involves the following hypothesis: The quality of an image could be evaluated by comparing it against a reference signal of perfect quality. A measure of the similarity between the reference image and the image being evaluated could be calibrated to serve as a measure of perceptual quality. A full-reference algorithm as proposed by Sheikh (2004) [7] computes the similarity between the images or video whose quality is to be evaluated (called the *test* signal) and the associated reference signal. The objective quality metrics commonly used to measure the perceived image quality are as given below:

- Root mean-squared-error (RMSE)
- Peak signal to noise ratio (PSNR)
- Mean structural similarity index measure (MSSIM)
- Similarity factor (SF)
- Block boundary measure (BBM)
-

3.1.1 Root mean-squared-error (RMSE):

One obvious way of measuring this similarity is to compute an error signal by subtracting the test signal from the reference, and then computing the average energy of the error signal. The root mean-squared-error (RMSE) as given in equation (2) by Gonzales and Woods (2003)[9] is the simplest, and the most widely used, FRQA method.

$$e_{rms} = \left[\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\hat{f}(x,y) - f(x,y)\right]^2\right]^{1/2} \quad (2)$$

where $\hat{f}(x,y)$ is a source image of size M×N pixels and $f(x,y)$ is processed image of size M×N pixels.

The RMSE as given by the above equation correlates poorly with subjective image quality.

3.1.2 Peak signal to noise ratio (PSNR):

The objective quality of the decompressed image is evaluated using the peak signal to noise ratio by Hsu and Chen (1993)[10], For N×N images with [0, 255] gray level range, PSNR is defined in db as:

$$\mathrm{PSNR} = 10 \times \log_{10}\frac{1}{N}\sum_{i=1}^{N}\frac{255^2}{(x_i - y_i)^2}\ \mathrm{db} \qquad (3)$$

where N is the number of samples and x_i and y_i are the gray levels of the original and reconstructed images respectively. It is well known that PSNR is not always a good measure to reflect the subjective image quality; even though it is one of the most popular criteria employed in image processing.

3.1.3.Mean structural similarity index measure (MSSIM):

In the last three decades, a great deal of effort has gone into the development of quality assessment methods that take advantage of the characteristics of the HVS. Many researchers have invested time in to the development of quality assessment methods that take advantage of known characteristics of the human visual system. Wang [11] proposed image metrics based on the assumption that the HVS is highly adapted to extract structural information from the viewing field. Wang et al. (2004) [11 proposed that Structural Similarity measure (SSIM) compares local patterns of pixel intensities after they have been normalized for luminance and contrast. The SSIM indices measure the structural similarity between two image signals. Suppose a and b are two non-negative image signals, if one of the signals is considered to have perfect quality, the similarity measure can be used as a quantitative measurement of the quality of the second signal and is computed as:

$$\mathrm{SSIM}(a,b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{ab} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)} \quad (4)$$

where μ_a , μ_b and σ_a, σ_b are mean intensities and standard deviations for two non-negative images a and b, respectively. C_1 and C_2 are constants. In discrete form σ_{ab} can be estimated as:

$$\sigma_{ab} = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - \mu_a)\,(b_i - \mu_b) \qquad (5)$$

For image quality assessment, it is useful to apply the SSIM index locally rather than globally. The local statistics are computed within a local $w \times w$ square window, which moves pixel-by-pixel over the entire image. At each step, the local statistics and SSIM index are calculated within the local window. In practice, a single overall quality measure of the entire image is required. The MSSIM index used to evaluate the overall image quality is computed as:

$$\mathrm{MSSIM}(A,B) = \frac{1}{W}\sum_{i=1}^{W} SSIM(a_i\,, b_i) \qquad (6)$$

where A and B are the original and reconstructed images respectively; a_i and b_i are the image contents at the ith local window; and W is the number of local windows of the image.

3.1.4 Block boundary measure (BBM):

Singh et al. (2007) [12]proposed a new index BBM called block boundary measure. This index is used to measure the quantum of blocking artifacts at block boundaries.
For the image f given by:

$$f = \begin{Bmatrix} p_{m,n}^{0,0} & p_{m,n}^{0,1} & \dots & p_{m,n}^{0,z-1} \\ p_{m,n}^{1,0} & p_{m,n}^{1,1} & \dots & p_{m,n}^{1,z-1} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m,n}^{z-1,0} & p_{m,n}^{z-1,1} \dots & & \dots p_{m,n}^{z-1,z-1} \end{Bmatrix} \qquad (7)$$

where $p_{m,n}^{a,b}$ represents (m, n)th pixel intensity value in (a, b)th block, z is the number of blocks in the image along horizontal or vertical directions. The block boundary measure between two horizontal 8×8 blocks called BBM_{ver} is computed as follows:

$$BBM_{ver} = \frac{1}{(z-1)^2}\sum_{a=0}^{z-2}\sum_{b=0}^{z-2}\left\|p_{m,7}^{a,b} - p_{m,0}^{a,b+1}\right\|, \quad \forall_m \in [0,7] \qquad (8)$$

Similarly, the BBM for the horizontal block boundaries can be obtained as:

$$BBM_{hor} = \frac{1}{(z-1)^2}\sum_{a=0}^{z-2}\sum_{b=0}^{z-2}\left\|p_{7,n}^{a,b} - p_{0,n}^{a,b+1}\right\|, \quad \forall_n \in [0,7] \qquad (9)$$

3.2 Objective No-Reference Quality Assessment

Providing the reference signal to the quality assessment algorithm renders such algorithms infeasible for most applications. FRQA methods are basically useful only as a `lab tool' for designing systems since the resource challenges in providing the reference signal in, say, a quality monitoring application, are virtually insurmountable. Thus, a different framework is needed for objective quality assessment, in which the algorithm does not rely on the availability of the reference signal to evaluate the quality of the test signal.

This is the so-called no-reference (NR) quality assessment problem. It is obvious that human observers can easily evaluate the quality of images or videos without needing to view the reference signal. Thus it is believed that the design of NRQA should theoretically be possible. Unfortunately, the NR problem is, as yet, an unsolved problem with no known generic NRQA algorithm that exists today. Only limited success has been achieved by limiting the scope of NR methods to specific distortion types as suggested by Sheikh (7).

3.3 Objective Reduced-Reference Quality Assessment

Reduced-reference quality assessment methods are those in which partial information regarding the reference image is available. Quality Assessment algorithms use this partial reference information to judge the quality of the distorted signal.

4 CONCLUSIONS

In this paper, an overview of image quality assessment was presented. Quality assessment is classified on the degree of human involvement. Subjective measures are by definition based on HVS and may therefore accurately relate to the perceived quality. In the past, simple objective measures of image quality were poorly correlated with perceived picture quality. Recently developed objective quality metrics, have better correlation with subjective measures.

REFERENCES

[1] S. Grgic, M. Grgic, and M. Mrak, "Reliability of objective picture quality measures," Journal of Electrical Engineering, vol. 55, issue 1-2, pp. 3-10, 2004.

[2] Y. Horita, M. Sato, Y. Kawayoke, Z. M. P. Sazzad, and K. Shibata, "Image Quality evaluation model based on local features and segmentation," in IEEE International Conference on Image Processing, Atlanta, *GA*, 2006, pp. 405-408.

[3] K. Tan, M. Ghanbari, and D. E. Pearson, "An objective measurement tool for MPEG video quality," Signal Processing, vol. 70, no. 3, pp. 279-294, 1998.

[4] N. Narita, "Subjective evaluation method for quality of coded images," IEEE Transactions on Broadcasting, vol. 40, pp. 7-13, 1994.

[5] A. C. Bovik, The Handbook of Video Databases: Design and Applications. USA, CRC Press, 2003

[6] M. Knee, "The picture appraisal rating (PAR) - a single-ended picture quality measure for MPEG-2," *in Proceedings of the International Broadcast Convention,* Amsterdam, 2000, pp. 1-6.

[7] H. R. Sheikh, "Image quality assessment using natural scene statistics," Ph.D. dissertation, University of Texas, Austin, May 2004.

[8] D. Zhang, "Information theoretic criteria for image quality assessment based on natural scene statistics," M.Sc. thesis, University of Waterloo Ontario, Canada, 2006.

[9] R. C. Gonzales, and R. E. Woods, Digital Image Processing. 3rd edition, New Delhi, Pearson Education, 2003.

[10] Y. F. Hsu, and Y. C, Chen, "A new adaptive separable median filter for removing blocking effects," IEEE Transactions on Consumer Electronics, vol. 39, issue 3, pp. 510-513, 1993.

[11] Z. Wang, A.C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, issue 4, pp. 600-612, 2004.

[12] S. Singh, V. Kumar, and H. K. Verma, "Reduction of blocking artifacts in JPEG compressed images," Journal Digital Signal Processing, vol. 17, issue 1, pp.225-243, 2007.