Competitive Analysis of Existing Image Quality Assessment Methods

Rushikesh Borse
PhD Research Scholar
Dept of Electrical Engineering
Indian Institute of Technology Bombay
Mumbai, India
rpborse@ec.iitb.ac.in
rpborse@gmail.com

Prerana Markad
Post Graduate Student
Dept of Electronics and Telecommunication Engg.
Sinhgad Academy of Engineering
Pune, India
preranaanu77@gmail.com

***Abstract*— Image quality assessment (IQA) has become indispensable part of image processing applications since times. Image quality assessment deals with the evaluation of image quality using different measures. The Goal of image quality assessment is to provide a metric that can estimate the perceived quality of an image instinctively. The assessment of IQA metric gives optimum result when human visual system is taken in to consideration. In this paper, we described the current standards of image quality measures. We also do competitive analysis of these metrics. It is important to analyze the performance of these metrics in a comparative setting and analyze the strengths and weaknesses of these methods. Comparison of available image quality metrics is critically important in deciding as to which metric is better for a particular application.**

***Index Terms*— Image quality assessment (IQA), subjective metric, objective metrics, human visual system (HVS)**

1. INTRODUCTION

In recent years, uses of digital images have increased in applications such as medical imaging, remote sensing, compression, etc. However, digital images are subjected to various distortions while processing. These distortions include linear frequency distortion, blocking artifacts, ringing effect in an image, blurring, noise, color errors, etc. These sources of corruption may occur during processing or while capturing the image. So, image quality assessment (IQA) is being a weighty facet in different applications such as image acquisition, transmission, compression, storage, communication, printing, restoration, displaying, painting, photography, film and image enhancement [1-3].

Image quality assessment does try to compute the visual quality of the image. At the present time a lot of effort has been given to the research of incorporation of human perception in image and video processing applications. The image quality assessment (IQA) research, therefore, aims at designing a metric for estimation of quality in a way that is consistent with human visual system. In order to design a better quality metric, incorporation of human visual features is substantive. All the proposed IQA metrics can be classified into two general classes: subjective measurement and objective measurement [4].

1.1 Subjective measurement

Subjective image quality measure involves use of human subjects in assessing the quality of an image. It is concerned with how the image is perceived by a human being. Based on the perception of quality of image, subject gives his or her opinion. In this method a number of observers are selected and asked to rate the quality of images. Range of the subjective scores used for the test is 1-11. Subjective measurement is accepted method for accurately evaluating the quality of images. Ultimately, it is a human being who is going to determine the quality of an image. So it is very reliable and accurate metric of quality assessment. However, subjective measure is time consuming, very inconvenient and can be expensive at times.

1.2 Objective Measurement

Objective measurement based metrics analyze images and calculate their quality without human involvement. Objective measurement can be further classified according to the availability of the original image, with which the distorted image is to be compared. If a complete original image is available it is called ‘full reference’. If original image is not available it is ‘no reference’ or ‘blind original. And, if original image is partially available in the form of extracted features it is called ‘reduced reference’. Objective measurement is easy to operate. It is optimum if we take HVS into consideration. This measure saves time as compared to subjective quality measurement [5], [6]. This paper does focus on full reference IQA metrics.

In this paper, we present results of well known IQA algorithms and compare their performance. We also do subjective analysis and support the results with objective analysis. The paper is organized as follows: Section 2 describes the existing methods of IQA and their limitations, Section 3 highlights the competitive analysis, and Section 4 concludes the paper.

2. IMAGE QUALITY ASSESSMENT METRICS

In this paper we discuss various existing IQA metrics such as Mean Square Error, Peak to Signal Noise Ratio, Average Difference, Maximum Difference, Mean Absolute Error,

978-1-4799-3080-7/14/$31.00

Normalized cross-correlation and Structural Content. In the following equations we use these notations: x(i,j) denotes the samples of original images and y(i,j) denotes distorted images. M and N are number of pixels in row and column respectively and (i,j) is the pixel position of M×N image.

2.1 Mean Squared Error (MSE):

Simplest and the most widely used objective IQA metric is MSE. This metric is frequently used in signal processing and is defined as follows [6]:

$$MSE = \frac{1}{MN}\sum_{1}^{M}\sum_{1}^{N}[x(i,j) - y(i,j)]^2 \quad (1)$$

MSE is zero when x(i,j) = y(i,j). If MSE is small enough, this corresponds to a high quality image. As MSE reduces to zero, the pixel-by-pixel matching of the images becomes perfect. The squaring of the differences dampens small differences between two pixels but penalizes the large ones.

2.2 Average Difference (AD):

AD is simply the average of difference between the original and distorted images. It is given by the following equation [6]:

$$AD = \frac{1}{MN}\sum_{1}^{M}\sum_{1}^{N} x(i,j) - y(i,j) \quad (2)$$

Ideal value of average difference should be zero for high quality test image.

2.3 Maximum Difference (MD):

MD is the maximum of the error signal that is difference between the original and distorted image. It is defined as follows [6]:

$$MD = MAX\left| x(i,j) - y(i,j)\right| \quad (3)$$

Large value of MD indicates test image is of poor quality. Ideally it should be zero.

2.4 Mean Absolute Error (MAE):

MAE is simple to calculate. It is average of absolute difference between the original image and distorted image. It is defined as follows [6]:

$$MAE = \frac{1}{MN}\sum_{1}^{M}\sum_{1}^{N}\left| x(i,j) - y(i,j)\right| \quad (4)$$

Ideally, mean absolute error should be zero for a good quality image.

2.5 Normalized Absolute Error (NAE):

Normalized absolute error is calculated by following equation:

$$NAE = \sum_{1}^{M}\sum_{1}^{N}\frac{\left| x(i,j) - y(i,j)\right|}{\sum_1^M \sum_1^N \left| x(i,j)\right|} \quad (5)$$

Greater value of NAE indicates low quality. Ideally it should be zero.

2.6 Peak Signal to Noise Ratio (PSNR):

PSNR is the most widely used IQA metric for the evaluation of image quality. The PSNR value approaches infinity as the MSE approaches zero. It is defined as follows [6]:

$$PSNR = 10\log_{10}\frac{L^2}{\sqrt{MSE}} \quad (6)$$

where L is the dynamic range of the pixel values.

PSNR is appealing because it is simple to evaluate, it has clear physical meanings, and it is mathematically convenient when optimization is considered. It is very fast and easy to implement. However, PSNR does not take human visual system into consideration. Greater value of PSNR indicates greater image similarity. Ideally it should be infinity.

2.7 Normalized Cross-Correlation (NK):

The similarity between two digital images can also be quantified using correlation function. The correlation between two images (cross-correlation) is a standard approach to feature detection. It measures the similarity between two images. It is given by the equation [6]:

$$NK = \frac{\sum_1^M \sum_1^N [x(i,j) \times y(i,j)]}{\sum_1^M \sum_1^N [x(i,j)]^2} \quad (7)$$

2.8 Structural Content (SC):

Structural content is also a correlation based measure. It measures the similarity between two images. SC is given by the equation [6]:

$$SC = \frac{\sum_1^M \sum_1^N [y(i,j)]^2}{\sum_1^M \sum_1^N [x(i,j)]^2} \quad (8)$$

Ideally, a good quality test image should have SC equal to one.

2.9 Structural Similarity Index Metric (SSIM):

Wang *et al.* developed SSIM metric for the evaluation of the image quality which is based on the hypothesis that the HVS is extremely susceptible to structural information in an image. This metric compares the local patterns of pixel values that have been normalized for luminance and contrast [7]. It is designed to provide improvement over conventional metrics like MSE and PSNR.

General formula of SSIM is as follows:

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \quad (9)$$

where $\alpha > 0$, β>0 and γ >0 are parameters used to adjust the relative importance of the three components; x, y are image patches and l(x,y), c(x,y), s(x,y) are defined as follows:

$$l(x,y) = \frac{2\bar{x}\bar{y} + C_1}{\bar{x}^2 + \bar{y}^2 + C_1} \quad (10)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (11)$$

$$s(x,y) = \frac{2\sigma_x\sigma_y + C_3}{\sigma_x^2 + \sigma_y^2 + C_3} \quad (12)$$

l(*x*,*y*) is luminance comparison, *c*(*x*,*y*) is contrast comparison, and *s*(*x*,*y*) is structural comparison; C_1, C_2, C_3 are constants. N is number of pixels. The mean and variance values: x□, y□, σ_x, σ_y and σ_{xy} are given as:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (13)$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (14)$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \quad (15)$$

$$\sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \quad (16)$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \quad (17)$$

$$SSIM = \frac{(2 \times \bar{x} \times \bar{y} + C_1)(2 \times \sigma_{xy} \times C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)[(\bar{x})^2 + (\bar{y})^2 + C_1]} \quad (18)$$

The overall image quality Mean SSIM is obtained by computing the average of SSIM values over all windows:

$$MSSIM = \frac{1}{M}\sum_{1}^{M} SSIM_j \quad (19)$$

Where M = total number of windows.

Unlike MSE and PSNR, SSIM is based on the assumption that the HVS is highly adaptive for extracting structural information from the image.

2.10 Universal Quality Index Metric (UQI):

In 2002, Wang and Bovik proposed this metric [8]. UQI has powerful capability of measuring structural distortion that take place during the process of the image degradation. It is given by following expression:

$$Q = \frac{4 \times \sigma_{xy} \times \bar{x} \times \bar{y}}{(\sigma_x^2 + \sigma_y^2) \times [(\bar{x})^2 + (\bar{y})^2]} \quad (20)$$

Where x□, y□, σ_x^2, σ_y^2 and σ_{xy} are given as

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (21)$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (22)$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \quad (23)$$

$$\sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \quad (24)$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \quad (25)$$

UQI metric can be written as a fusion of three factors, correlation, luminance and contrast as shown in following equation. These components represent loss of correlation, luminance distortion and contrast distortion.

$$Q = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y} \times \frac{2 \times \bar{x} \times \bar{y}}{[(\bar{x})^2 + (\bar{y})^2]} \times \frac{2 \times \sigma_x \times \sigma_y}{(\sigma_x^2 + \sigma_y^2)} \quad (26)$$

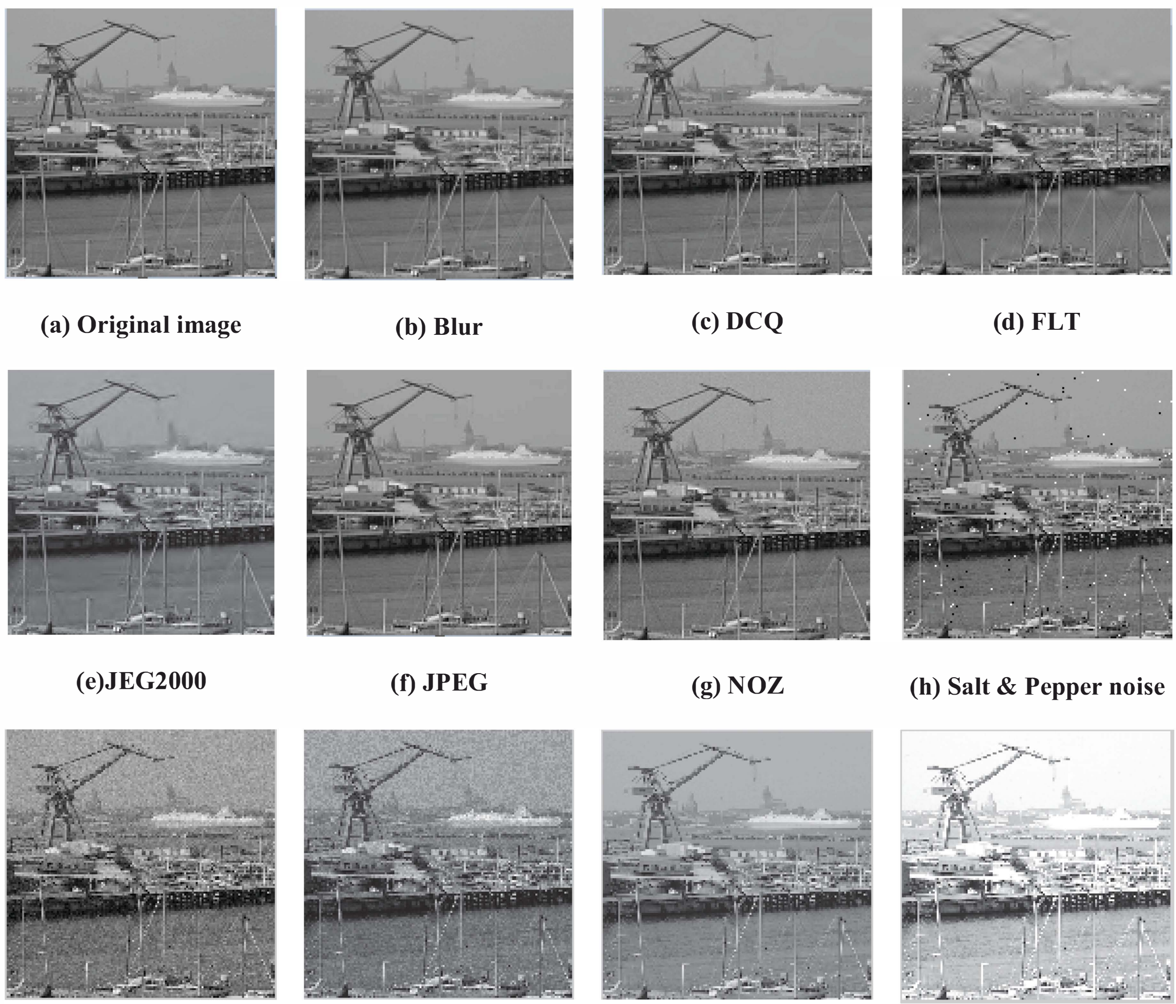
3. COMPETITIVE ANALYSIS

For evaluation of available quality measures we use different types of distortions listed in table 1. We perform the analysis of these metrics for ‘harbour.bmp’ image. The harbor image is taken from A57 database [9]. The original “harbor.bmp” image is used as reference image. The distortions used in fig1(b) to fig1(g) are from database A57: Gaussian blur, JPEG-2000+DCQ compression, flat allocation (equal distortion at all scales), baseline JPEG-2000 compression, baseline JPEG compression, and Gaussian white noise respectively. Along with the set of distorted images given in database, we carry different distortions to have more distorted images. Some of these are frequency domain distortions, while others are spatial domain distortions. Various types of distortion used are Logarithm transform, Gamma transform, Salt & pepper noise, Gaussian noise and Speckle noise. We calculate the IQA of fig1(b) to fig1(l), which are presented in table1. We also do compare the values of IQA evaluated with all metrics so as to decide which metric is suitable particular application. For the subjective analysis, 25 observers were selected and asked to rate the images shown in fig. 1. The average of the scores is calculated for each image, shown in table 1.

The assessment of the corrupted images as shown in figure 1 is described in table 1. Subjective score of each image is also presented. It can be clearly seen from the evaluation of current metrics, that none of these metrics support human visual system (HVS). Image (c) looks better than all other images. It also has got highest subjective score. Even though it’s MSE is not minimum. MSE for images (h), (k) and (l) are lesser than that of (c). Also MSE values of (b), (d), (e), (f) and

(g) are greater than (h), (k) and (l) even though they are visually better. Same is the case with AD. AD of (c) is greater than (h), (k) and (l). Visually worst image (i) should have maximum MD, but it has lesser MD than (c)-(h). Ideally MAE value should be least for (c), but (k) and (l) have got lesser values than (c). Image (k) got zero NAE value, still (c) has non zero NAE. Also (l) has less NAE value than (c). Ideally, PSNR value should be infinity and practically as high as possible. Here, (k) is visually better than (d), (e), (g)-(j) and (l) still has lesser PSNR. Image (l) whose subjective score equals five, has got highest PSNR value. Image (b) is visually better than (d) and (j) still has smaller value of NK and SC. Next parameter is SSIM. Ideally visually worst image should have least SSIM value. Here, image (k) whose subjective score is eight has got least SSIM. Image (l) has lesser UQI value than (d), (h)-(j) even though it has greater subjective score.

As discussed, all the parameters do not evaluate image quality in accordance with human perception. Still, Parameters like SSIM, PSNR and AD are better than other metrics like MSE, NAE, NK, SC, MAE. SSIM is expected to be 1 for good quality image. Here, image (c) is visually better than all other images. Its SSIM value is higher than rest images and near to one. Also image (c) has higher PSNR value. Ideally PSNR should be infinite and image (c) has second highest PSNR of all.



(a) Original image **(b) Blur** **(c) DCQ** **(d) FLT**

(e)JEG2000 **(f) JPEG** **(g) NOZ** **(h) Salt & Pepper noise**

(i) Gaussian noise **(j) Speckle noise** **(k) Gamma Transform** **(l) Logarithm transform**

Fig. 1: (a) A reference image; (b) to (f) are the mid contrast (0.145 total contrast) distorted versions of (a) in the A57 database. Distortion types of (b) to (f) are *Gaussian blur, JPEG-2000+DCQ compression*, *Flat allocation*, *Baseline JPEG-2000 compression*, *Baseline JPEG compression*, and Gaussian white noise, respectively; (h) to (l) are distorted images with distortion types Salt & Pepper noise, Gaussian noise, Speckle noise, Gamma transform and Logarithm transform respectively.

Table 1: Competitive analysis of quality measures of harbor image with different types of distortions.

Metrics	Fig 1(b)	Fig 1(c)	Fig 1(d)	Fig 1(e)	Fig 1(f)	Fig 1(g)	Fig 1(h)	Fig 1(i)	Fig 1(j)	Fig 1(k)	Fig 1(l)
Subjective Score	9	11	3	7	10	6	4	1	2	8	5
MSE	37.095	22.289	24.772	28.900	26.607	32.440	0.6294	47.923	40.532	0	0.0926
AD	3.7141	2.5410	2.8473	3.1215	2.9724	3.3572	0.3116	4.5426	3.8546	0	-0.2975
MD	43	78	92	79	91	70	225	52	28	-4	-0.0590
MAE	3.7141	2.5410	2.8473	3.1215	2.9724	3.3572	0.3116	4.5426	3.8546	0	0.2975
NAE	0.0295	0.0202	0.0226	0.0248	0.0236	0.0266	0.0025	0.0361	0.0306	0	0.6020
PSNR	28.7580	28.9421	28.8916	28.8256	28.9435	28.7738	28.5055	27.0258	28.7532	23.0889	58.4973
NK	0.9988	1	0.9996	0.9993	0.9991	0.9983	0.9975	0.9981	1	1	1.5841
SC	1.0029	1	1.0008	1.0013	1.0017	1.0034	1.0025	1.0040	1.0001	1	0.3978
SSIM	0.9996	0.9999	0.9998	0.9991	0.9999	0.9942	0.9996	0.9994	0.9998	0.9734	0.9878
UQI	0.9998	0.9999	0.9999	0.9996	0.9999	0.9961	0.9998	0.9997	0.9999	0.9745	0.8524

4. CONCLUSION

The purpose of this paper is to understand fundamental approaches of image quality assessment methods. In this paper, we attempt to do competitive analysis of these existing image quality metrics. We provide a reasonable discussion about each metric in section 3. We conclude that structural based approach SSIM outperforms other approaches.

The important point to be considered in the design of image quality assessment algorithm is human visual system (HVS). So, there is an urgent need of image quality metric which better correlates with human perception. We finally settle at point that SSIM is better representation of HVS amongst the other factors and we wish to propose improvements to it.

REFERENCES

[1] Z. Wang, A.C. Bovik, Modern Image Quality Assessment. San Rafael, CA: Morgan and Claypool Publishing Company, New York, 2006.

[2] I. Avcibas¸, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imag.*, vol. 11, no. 2, pp. 206–23, Apr. 2002.

[3] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, pp. 3441-3452, November 2006.

[4] A. B. Watson, "Visual detection of spatial contrast patterns: Evaluation of five simple models," *Optics Express*, vol. 6, pp. 12–33, Jan. 2000.

[5] *Methodology for the Subjective Assessment of the Quality for Television Pictures*, ITU-R Rec. BT. 500-11.

[6] M. Eskicioglu and P. S. Fisher, "Image Quality Measures and Their Performance," *IEEE Transactions on Communications*, vol. 43, no. 12, December 1995, pp. 2959-2965.

[7] Z. Wang A C. Bovik, H. R. Sheikh, and Simoncelli, "Image quality assessment: From error visibility to structural similarity", IEEE Trans. Image Processing, vo. 13, no. 4, pp. 600-612, Apr. 2004.

[8] Z. Wang and A. C. Bovik, " Universal Image Quality Index" *IEEE SP letters*, vol. 9, pp. 81-84, March 2002.

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