Comparison of Image Quality Assessment: PSNR, HVS, SSIM, UIQI

Yusra A. Y. Al-Najjar, Dr. Der Chen Soong

Abstract — Measuring the quality of the image is a complicated and hard process since humans opinion is affected by physical and psychological parameters. Many techniques are proposed for measuring the quality of the image but none of it is considered to be perfect for measuring the quality. Image quality assessment plays an important role in the field of image processing. Many studies have been done on image quality measurements based on different techniques such as pixel-difference, correlation, edge detection, neural networks (NN), region of interest(ROI), human visual system (HVS). The good IQM must be accurate and consistent in predicting the quality. Most IQ metrics are related to the difference between two images (the original and the distorted image).

Index Terms — HVS, Image Quality Assessment, MSE, PSNR, UIQI

1 Introduction

THE quality of the image degrades from the minute it is captured to the time it is displayed to the human observer. The image is subject to many kinds of distortions during the stages that it might pass through such as storing, processing, compressing, and transmitting, etc...

In evaluating image quality there are two followed methods, the subjective and the objective method. The subjective method evaluation is considered costly, expensive, and time consuming; since we have to select a number of observers, show then a number of images and ask them to score images quality depending on their own opinion. The objective evaluation uses automatic algorithms to assess the quality of the image without human interfere.

Objective image quality matrices are divided into different categories depending on the existence of the original image:

- Full-reference: where the reference image is available
- Reduced-reference: where the reference image exists partially in a set of extracted features as information that helps in the evaluation
- No-reference: where there is no reference image. This is also called "blind quality assessment".

This paper will refer to full-reference image quality matrices.Mean Opinion Score (MOS) is the well-known approach for subjective image quality assessment. In this approach a group of people are asked to compare original images with distorted images in order to estimate the quality of the distorted image. The mean score is taken as the image quality index. Despite that this process reflects human perception, it is considered time consuming and unpractical to use in conjunction with other image processing algorithms. That is why we need a strong metric to correlate with subjective assessment.

2 Procedure for Paper Submission

2.1 Background

Image quality assessment points to measure the degradation in digital images in order to improve the quality of the resultant image. In practice we have two kinds of evaluation subjective and objective evaluation [1]. Subjective evaluation is inconvenient, time-consuming and expensive.

Lately, lots of efforts have been done to develop objective image quality metrics. MSE, PSNR, and SSIM are the most commonly used objective image quality measures. In this paper we concentrate on full reference objective quality metric.

In our experiment we have both the original and the distorted images and focus on full-reference quality measures. Full reference image quality measures could be classified into six classes of objective image assessment measures [2], that is:

1) Pixel difference-based measures: The mean square error (MSE), signal-to-noise ratio (SNR) and peak signal-to-noise ratio (PSNR). These measures are easy to evaluate
2) Correlation-based measures: Correlation is used to measure the difference between two digital images. In image quality assessment, correlation of pixels is used as a measure of the image quality.
3) Edge-based measure: In this class the edges in the original and the distorted images are found, then a measure of displacement of edge positions or there consistency are used to find the image quality for the whole image.
4) Spectral distance-based measures: Discrete Fourier Transform is applied on the original and the

- Al-Najjar Y. Author is a PhD student in UniversitiTenagaNasional, Malaysia (e-mail: yusra.najjar.2011@gmail.com).
- Soong Der Chen is with the College of IT, UniversitiTenagaNasional, Putrajaya, Malaysia (e-mail: chensoong@uniten.edu.my).

distorted images. The difference of the Fourier magnitude or phase spectral is used as a measure of image quality.

5) Context-based measures: Instead of comparing pixels in original and distorted images, pixel neighborhoods are compared against each other by finding the multidimensional context probability to use it for measuring image quality.
6) Human Visual System-based measures (HVS): Here image quality is measured as the human eye would do. Humans usually use contrast, color, and frequency changes in their measures.

Researches in the field of full reference HVS are related to understanding human visual perception, where the quality is computed by comparing it against a reference image. The first five types of image quality metrics are so called simple statistics error metrics while the last one is called feature based metric.

2.2. Examples of Image Quality Metrics

1. Pixel Difference Measurement

Types related to this category are like MSE and PSNR:

i) Mean Square Error (MSE), MSE is computed by averaging the squared intensity of the original (input) image and the resultant (output) image pixels as in (1).

$$MSE = \frac{1}{NM}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} e(m,n)^2 \qquad (1)$$

Where e(m, n) is the error difference between the original and the distorted images.

ii) Peak Signal-to-Noise Ratio (PSNR), Signal–to-noise ratio (SNR) is a mathematical measure of image quality based on the pixel difference between two images [3]. The SNR measure is an estimate of quality of reconstructed image compared with original image. PSNR is defined as in (2)

$$PSNR = 10log\frac{s^2}{MSE} \qquad (2)$$

where s = 255 for an 8-bit image. The PSNR is basically the SNR when all pixel values are equal to the maximum possible value.

2. Human Visual Based Measurement

Human Visual System (HVS) is another approach of measuring image quality [4], [5], [6], [7]. The HVS is a method that uses human eye as a reference. The main idea is that humans are interested in different attributes of the image other than taking it as a whole. These attributes include brightness, contrast, texture, orientation…etc.

Despite that HVS measurement is very complex to be understood with psychophysical means, HVS is the tool for human being to understand the world surrounding and the tool that reveals brain secrets. A large number of physiological [8] and psychophysical [9] experiments show physiological mark and are the only way to understand the phenomenon.

Both images – original and distorted– are transformed into frequency domain. Two techniques are normally used to transform the images into the frequency domain, Discrete Fourier Transform (DFT), and Wavelet Transform. After transforming images into frequency domain, a band-pass filter known as Contrast Sensitivity Function (CSF), is applied to the original and the distorted images. The CSF has a band-pass characteristic which correlates with how human eye scale an image in the frequency domain. A band filter in the frequency domain can be define as in (3)

$$H(w) = \begin{cases} 0.05e^{w^{0.554}} & w < 7 \\ e^{-9[|\log_{10}w - \log_{10}9|]^{2.3}} & w \geq 7 \end{cases} \qquad (3)$$

Where $w = (u^2 + v^2)^{1/2}$ and u and v are the spatial frequencies. After the band pass filter is used, many different approaches could be used to measure image quality, the simplest approach is use of the MSE on the processed images.

Many different types of HVS models have been developed to measure image quality, however, among all objective measures, the HVS measure is considered to be the closest to subjective measures. Below are listed two human visual system based metrics:

a. Universal Image Quality Index (UIQI)
b. Structural Similarity Index (SSIM),

a) Universal Image Quality Index

In 2002, Wang and Bovik proposed this measure[10], it breaks the comparison between original and distorted image into three comparisons: luminance, contrast, and structural comparisons as in (4), (5), and (6).

$$l(x,y) = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2} \qquad (4)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \qquad (5)$$

$$s(x,y) = \frac{2\sigma_{xy}}{\sigma_x + \sigma_y} \qquad (6)$$

Where $\mu_x\mu_y$ denotes the mean values of original and distorted images. And $\sigma_x\sigma_y$ denotes the standard deviation of original and distorted images, and σ_{xy} is the covariance of both images.

Based on the above three comparisons the UIQI is given in (7)

ISSN 2229-5518

$$UIQI(x,y) = l(x,y).c(x,y).s(x,y) = \frac{4\mu_x\mu_y\mu_{xy}}{(\mu_x^2+\mu_y^2)(\sigma_x^2+\sigma_y^2)} \quad (7)$$

UIQI is a simple measure that counts only on first and second order statistic of the original and distorted images.

UIQI is considered unstable measure and doesn't correlate will with subjective assessment that is why Wang et. al proposed structural similarity index metric.

b) Structural Similarity Index

Wang et. al[10], proposed Structural Similarity Index as an improvement for UIQI, as in Fig.1.

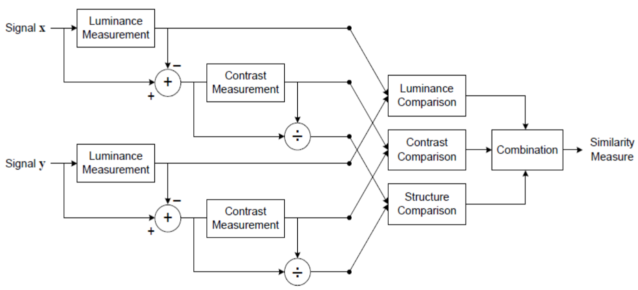

Fig.1: Diagram of Structural Similarity Index (SSIM) measurement system

The mean structural similarity index is computed as follows:

Firstly, the original and distorted images are divided into blocks of size 8 x 8 and then the blocks are converted into vectors. Secondly, two means and two standard derivations and one covariance value are computed from the imagesas in (8), (9), and (10).

$$\mu_x = \frac{1}{T}\sum_{i=1}^{T} x_i \qquad \mu_y = \frac{1}{T}\sum_{i=1}^{T} y_i \quad (8)$$

$$\sigma_x^2 = \frac{1}{T-1}\sum_{i=1}^{T}(x_i-\bar{x})^2 \; \sigma_y^2 = \frac{1}{T-1}\sum_{i=1}^{T}(y_i-\bar{y})^2 \quad (9)$$

$$\sigma_{xy}^2 = \frac{1}{T-1}\sum_{i=1}^{T}(x_i-\bar{x})(y_i-\bar{y}) \quad (10)$$

Thirdly, luminance, contrast, and structure comparisons based on statistical values are computed like in UIQI, the structural similarity index measure between images x and y is given by (11).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (11)$$

Where c_1 and c_2 are constants.

Like in UIQI, SSIM is applied locally using sliding window of size B x B that moves pixel by pixel horizontally and vertically covering all the rows and columns of the image, starting from top-left corner of the image. The overall image quality MSSIM is obtained by calculating the mean of SSIM values over all windows as in (12):

$$MSSIM = \frac{1}{P}\sum_{j=1}^{P} SSIM_j \quad (12)$$

where p is the number of sliding windows.

UIQI and SSIM are more accurate and consistence than MSE and PSNR despite they cost more.

3 METHODOLOGY

First selected images were converted into gray images using the function RGB2grayin MATLAB, then the metrics were implemented upon these images and last a comparison has been done between four objective evaluations: pixel-difference based measurement Peak Signal-to-Noise Ratio (PSNR), HVS using Fourier Transform, Structural Similarity Index (SSIM), and Universal Image Quality Index (UIQI) metrics by simulating them using MATLAB software.

MATLAB software is identical for dealing with graphics since it has an image processing tool box, beside it got lots of built in math function that can help in evaluating many statistics.

This comparison included the following nine original and distorted images, hats, ship, window, toys, house, butterfly, parrots, airplane, and light house selected from "Lossless True Color Image Suite" provided by "LIVE Image Quality Assessment Database" provided by Laboratory of Image and Video Engineering at University of Texas, Austin.

ISSN 2229-5518

Fig.2: The nine images from (a) to (i) used in the experiment

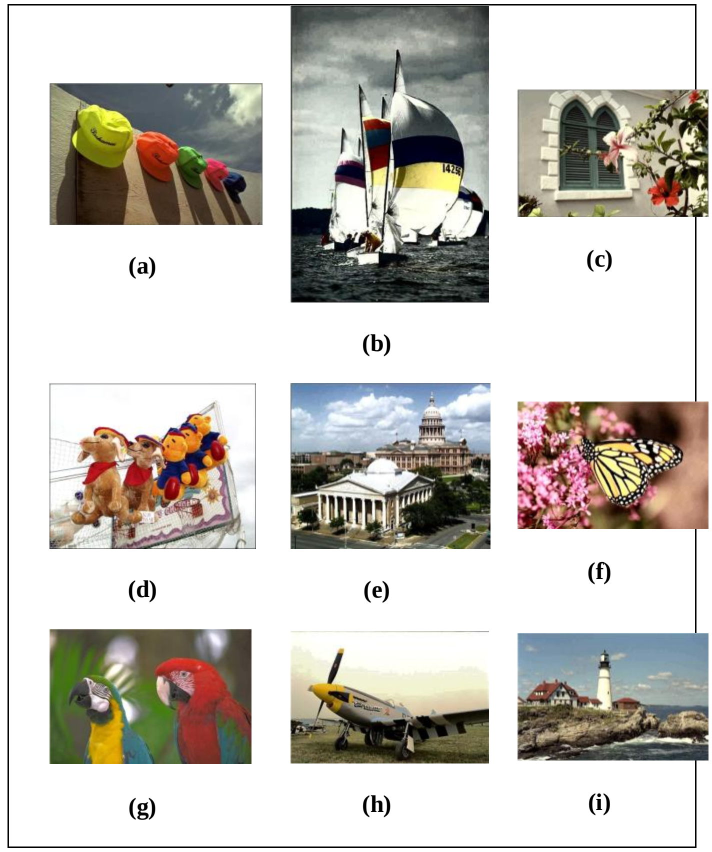

Fig.3:Distorted images from (a) to (i) used in the experiment

4 Results

All used Image quality metrics are objective measurements that are automatics and mathematical defined algorithms.

After applying some distortion (contrast enhancement) to the original nine images, see Fig.2, we got the distorted nine images included here, see Fig.3, and the image quality is applied to theses distorted images and the results are compared.

Measuring image quality for the nine images gave the results included in TABLE I.

TABLEI

Comparison of PSNR, SSIM, HVS, and UIQI Image Quality Measurements Applied on Images Included in the Experiment

Image	PSNR	SSIM	HVS	UIQI
Hats	18.4	0.78	0.34	0.88
Ship	12.0	0.55	2.68	0.76
Window	13.1	0.57	1.04	0.83
Toys	15.17	0.86	0.35	0.93
House	13.67	0.59	1.9	0.83
Butterfly	20.5	0.9	0.14	0.95

Airplane	15.72	0.76	0.53	0.86
Light house	17.18	0.71	1.24	0.92

5 Discussion

It can be seen from TABLEI in previous section that different types of Image Quality metrics differ in value according to types of distortion in the image and that it is hard to get the same quality value even if the same distortion is implemented on different images.Despite that SSIM was built from UIQI, it is noticed that the result given by UIQI is closer to 1 than SSIM. From previous discussion we can see that we need to work more to get close to subjective image quality measurement.

6 Conclusion

There are many different types of image quality metrics implemented for getting the quality of an image, but there are still limitations. Despite subjective IQM are time consuming. And expensive but it still do better than objective IQM, and the objective IQM field is still open and need lots of work to co-operate with subjective IQM.

7 References

[1] C. A. B. R. H. S. a. P. E. S. Z. Wang, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

[2] I. A. ,. B. S. ,. K. S. Ismail Avcibas, "Statistical Evaluation of Quality Measures in Image Quality Compression," Journal of Electronic Imaging, 2002.

[3] a. L. M. Jean-Bernard Martens, "Image dissimilarity," Signal Processing, vol. 70, no. 3, pp. 155-176, 1998.

[4] W. a. B. Lagendijk R.L., "Visual image quality model_Tech," Delft University of Technology. The Netherlands, 1995.

[5] T. D. Kite, Design quality assessment of forward and inverse error difusion halftoning algorithm, Ph. D. thesis, University of Texas at Austin, 1998.

[6] P. M., "The Medical Image Display and Analysis Group at the North Carolina: Reminiscences and Philosophy," Computer Journal of IEEE Transactions on Medical Image, vol. 22, no. no.1, pp. 2-10, 2003.

[7] O. T. P. a. S. R. L., "Human visual system based wavelet decomposition for image compression," Journal of Visual Communication and Image Representation, vol. 6, no. no. 2, pp. 109-121, 1995.

[8] R. N. Carlson, "Foundation of Physiological Psycology (6th Edition)," Boston: Allyn & Bacon, 2004.

[9] D. A. C. J. E. B. Thomas T. Norton, Psychophysical Measurement of Visual Function, Boston: Butterworth-Heinemann, 2002.

[10] a. A. C. B. Zhou Wang, "A Universal Image Quality Index," IEEE SIGNAL PROCESSING LETTERS, vol. 9, pp. 81-84, 2002.

Yusra A. Y. Al-Najjarwas born on January 12, 1965 in Zarqa, Jordan. She received her B.E. degree in Computer Science from Kuwait University at 1986, M. Scin Computer Science from Al-Balqa'a Applied University at 2008. She is a LECTURER in KSA King Faisal University. Here research interest includes image enhancement, noise reduction, and Networks.

Soong Der Chen (M'2002) was born in October 6, 1973 in Kuala Lumpur, Malaysia. He received his B.E. in Electronics/Computer (1997), M.Sc (2000) and Ph.D (2008) from the Dept. of Comp. and Comm. Engineering in Universiti Putra Malaysia (UPM), Malaysia. Currently he is a senior lecturer in the Dept. of Graphics and Multimedia, College of IT, UniversitiTenagaNasional (UNITEN). His research interest includes image quality assessment, image enhancement, computer vision and image compression.