International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Comparative Analysis of Image Quality Assessment Metrics: MSE, PSNR, SSIM and FSIM

Yusra Al Najjar

Taibah University, Computer Science, AL Madinah, Saudi Arabia
Email: *yalnajar[at]taibahu.edu.sa*
DOI: https: //orcid. org/0000 - 0002 - 3369 - 4999

Abstract: ***Evaluating the quality of an image proves to be a multifaceted and intricate endeavor, given the nuanced nature of human perception influenced by an array of physical and psychological factors. Despite numerous proposed techniques aimed at measuring image quality, none emerge as flawless or universally applicable. In the realm of image processing, where precision is paramount, extensive research has explored diverse methodologies including point difference analysis, image correlation, edge detection, neural networks (NN), region of interest (ROI) analysis, and consideration of the human visual system (HVS). The essence of an effective image quality measure lies in its ability to furnish accurate and consistent predictions. This study concentrates on the full - reference image quality, which involves utilizing a reference image for comparison purposes. Additionally, the study directs its focus toward contrasting a range of comparison tools, each leveraging different metrics. Among the comparison tools scrutinized are those reliant on point - based measurements such as the mean square error (MSE) and the peak signal - to - noise ratio (PSNR). Conversely, others focus on the compositional aspects of images, exemplified by metrics like the Feature Index Matrix (FSIM) and the Structured Similarity Index Matrix (SSIM).***

Keywords: FSIM, full - reference, IQM, PSNR, SSIM

1. Introduction

The quality of an image deteriorates from the moment of capture to its presentation to a human observer. Degradation of perceived images is measured by image quality assessment methods.

Degradation manifests across several stages including storage, processing, compression, and transmission. The quality of an image is influenced by noise, contingent upon its correlation with the information sought by the viewer within the image.

Image quality assessment typically involves two methods: subjective and objective. Subjective evaluation, while accurate, is often perceived as expensive and time - consuming due to the necessity of selecting observers who must score image quality based on personal opinions. Objective evaluation relies on automatic algorithms to gauge image quality without human intervention. Objective image quality metrics are categorized based on the availability of the original image regarding reference [1]:

a) Full Reference (assessing the quality of an image with the reference image, where the reference image is considered the original image with good quality),
b) Reduced Reference (with partial reference information extracted from the original image,
c) No - Reference (there is no reference image, also known as "blind quality assessment").

The Mean Opinion Score (MOS) is a prominent method for subjective image quality assessment, where individuals compare original and distorted images to estimate the quality of the latter. The average score serves as the image quality index. Despite its reflection of human perception, this process is time - consuming and impractical to use alongside other image processing algorithms. Therefore, a robust metric is needed to align with subjective assessment regarding reference [2].

2. Literature Survey

Two types of evaluation methods are employed for image quality assessment: subjective and objective evaluation. Subjective evaluation, though accurate, is cumbersome, time - intensive, and costly. Consequently, significant efforts have been invested in developing objective image quality metrics. MSE, PSNR, and SSIM represent some of the most frequently utilized objective image quality measures. The Structural Similarity Index (SSIM) is a method used to measure the similarity between two images. SSIM considers the luminance, contrast, and structure of the images. Feature similarity index (FSIM) is related to phase congruency (PC) and gradient magnitude (GM).

This paper specifically searches into full - reference objective quality metrics addressing SSIM, FSIM, MSE, and PSNR.

Image quality metrics examples:

1) Pixel Difference Measurement
Types related to this category are MSE regarding reference [3] and PSNR regarding reference [4]:

a) Mean Square Error (MSE) is the most commonly used metric in image quality measure metrics, it is a full reference metric that quantifies the average square difference between pixels of the original and the pixels of the processed (distorted) image. The lower MSE value indicates it is closer to the original image meaning that there is less distortion or error between the original and the degraded image. MSE provides a numerical measure of discrepancy between original and degraded images, but it does not always correlate with human perception of image quality. MSE is computed by averaging the squared intensity of the original (input) image and the resultant (output) image pixels as in (1).

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (I_i - J_i)^{2} \quad \text{equation (1)}$$

Where:
- N is the total number of pixels in the images,
- I^i represents the intensity (pixel value) of the i^{th} pixel in the 1^{st} image,
- J^i represents the intensity of the corresponding pixel in the 2^{nd} image,

The term $(I_i – J_i)$ 2 computes the squared difference between the pixel values of the two images.

b) 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. It is a widely used metric for evaluating the quality of compressed images compared to original images. The SNR measures the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. PSNR is calculated using the mean squared error (MSE) between the original image and the degraded image. It is expressed in decibels (dB) and is defined as in (2):

$$PSNR = 10.log_{10}\left(\frac{MAX^2}{MSE}\right) \quad \text{equation (2)}$$

where MAX is the maximum possible pixel value (255 for an 8 - bit image). MSE is the mean squared error between the original and the distorted image.

PSNR is commonly used in image and video compression techniques. It assesses the effectiveness of compression algorithms by measuring how well the degraded image approximates the original. PSNR does not always correlate with human perception of image quality, in addition, PSNR is sensitive to small changes in pixel values and cannot capture subtle perceptual differences accurately.

2) Human Visual Based Measurements:

a) Human Visual System:
Human Visual System (HVS) is another approach to measuring image quality regarding reference [5]. The HVS is a method that uses the human eye as a reference. The main idea is that humans are interested in different attributes of the image rather 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 a tool for human beings to understand the world surrounding them and the tool that reveals brain secrets. Many physiological and psychophysical experiments show physiological marks and are the only way to understand the phenomenon. Both images – original and distorted– are transformed into the 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 a 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 that correlates with how the human eye scales an image in the frequency domain. A band filter in the frequency domain can be defined, regarding reference [6].

b) Universal Image Quality Index
In 2002, Wang and Bovik proposed this measure, it breaks the comparison between original and distorted images into three comparisons:luminance, contrast, and structural comparisons, regarding reference [7].

UIQI is a metric that evaluates the quality of an image based on the similarities between the local structure in the original image and the degraded image. It provides a single value that quantifies the quality of the distorted image compared to the original image. It is computed based on the mean and the standard deviation of the pixel intensities in local regions of the images. Equation (3) displays the formula for UIQI.

$$UIQI(x,y) = l(x,y).c(x,y).s(x,y) = \frac{4\mu_x\mu_y\mu_{xy}}{(\mu_x^2+\mu_y^2)\,(\sigma_x^2+\sigma_y^2)} \quad \text{equation (3)}$$

UIQI is considered an unstable measure and doesn’t correlate well with the subjective assessment which is why Wang et. al proposed the structural similarity index metric.

c) Structural Similarity Index (SSIM):
Wang et. al regarding reference [7], proposed Structural Similarity Index as an improvement for UIQI.

SSIM is a widely used metric for evaluating the similarity between two images. It assesses the structural similarity by comparing luminance, contrast, and structure. SSIM provides a measure of similarity between images as a value between - 1 and 1, where 1 indicates perfect similarity. So, it considers both luminance and structural similarities, regarding reference [8].

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 two standard derivations, and one covariance value are computed from the images, regarding reference8. Equation (4) displays the formula for SSIM, regarding reference [9].

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)\,(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)\,(\sigma_x^2+\sigma_y^2+c_2)} \quad \text{equation (4)}$$

While both UIQI and SSIM are used to assess image quality, SSIM is more widely adopted due to its comprehensive evaluation of structural similarity. SSIM considers perceptual aspects of image quality and provides a more detailed understanding of the similarity between images. UIQI and SSIM are more accurate and consistent than MSE and PSNR despite they cost more.

d) Featured Similarity Index Measure (FSIM):
The FSIM (Feature Similarity Index Measure) is an extension of SSIM that incorporates additional features of SSIM such as gradient magnitude and angle differences. FSIM aims to provide a more comprehensive measure of similarity between images by considering structural information beyond what SSIM offers. FSIM typically involves more complex computations and considers a wider range of features, making

Paper ID: SR24302013533 DOI: https://dx.doi.org/10.21275/SR24302013533

it potentially more accurate in assessing image quality, regarding reference [10].

To describe FSIM correctly, we need to describe two more criteria: Phase Congruency (PC) and Gradient Magnitude (GM).

- Phase congruency (PC): PC introduces a new approach for detecting image features. An essential characteristic of phase congruency is its immunity to variations in light within images. Moreover, it can detect a broader array of interesting features. It emphasizes image attributes within the frequency domain and remains invariant to changes in contrast.
- Gradient magnitude (GM): GM computation stands as a foundational aspect of digital image processing. Convolution masks represent the operators for calculating gradients. Various convolutional masks are available for gradient measurement. If (f (x)) represents an image and (Gx), and (Gy) denote its horizontal and vertical gradients, respectively, then the gradient magnitude of (f (x)) can be expressed as in 5.

$$\sqrt{G_x^2 + G_y^2} \qquad \text{equation (5)}$$

In this paper, we aim to evaluate the quality of images by measuring the similarity between them. Consider two images f1 (distorted image) and f2 (original image) and their respective phase congruency denoted as PC1 and PC2. The steps followed to compute FSIM are as follows:

1) Feature Extraction: extract relevant features from both the original and the distorted images. The Phase Congruency maps extracted from f1 and f2, along with the Gradient Magnitude (GM) maps G1 and G2 extracted from the same images, are utilized to define and compute the Feature Similarity Index Measure (the similarity between these two images can be determined by calculating equation, as in 6:

$$S_{\text{PC}} = \frac{2PC_1PC_2+T_1}{PC_1^2+PC_2^2+T_1} \qquad \text{equation (6)}$$

T1, a positive constant, serves to enhance the stability of Spc. Typically, T1 can be computed using the PC values. The provided equation delineates between two positive real numbers, with the scope confined to the range of 0 to 1.

2) Feature Comparison: compare the extracted features between the original and the distorted images. Similarly, to calculate the similarity between G1 and G2 as in 7:

$$S_{\text{G}} = \frac{2G_1G_2+T_2}{G_1^2+G_2^2+T_2} \qquad \text{equation (7)}$$

T2 represents a positive constant that relies on the dynamic range of gradient magnitude values. In this paper T1 and T2 remain constants, ensuring the convenient application of FSIM.

3) Combination of Features: combine the similarity measures obtained from different features into a single overall similarity measure. Now SPC and SG are combined to calculate the similarity SL of f1 and f2. S1 can be defined as in 8:

$$S_L(x) = [S_{PC}(x)]^{\propto}.\ [S_G(x)]^{\beta} \qquad \text{equation (8)}$$

Where parameters α and β are used to adjust the relative importance of PC and GM features.

4) Normalization and Scaling: Normalize the overall similarity measure to ensure that it falls within a predefined range, typically between 0 and 1, where 1 indicates perfect similarity.

While SSIM and FSIM are used for image quality assessment, FSIM offers a more advanced evaluation by considering additional features beyond luminance, contrast, and structure. However, FSIM may also involve higher computational complexity compared to SSIM [11].

3. Methodology

In our methodology, the image goes through a series of enhancement processes, including:

1) Convert the image to grayscale using the convert ('L') method, where 'L' stands for luminance using the Python Imaging Library (PIL), which is known as the Pillow library.
2) Transform into a blurred image using Python using various libraries such as OpenCV, the “GaussianBlur” function. The kernel size used is ((5, 5), 0) to determine the amount of blur, and 0 is the standard deviation of the Gaussien kernel which determines the amount of blur.
3) Subsequently, other metrics were applied to the image, followed by a comparison of four objective evaluations:
   - Mean Square Error (MSE)
   - Peak Signal - to - Noise Ratio (PSNR) relying on pixel differences,
   - Structural Similarity Index (SSIM),
   - and Featured Similarity Index (FSIM) metrics.

These assessments were simulated using Python software, which is adept at graphic processing due to its comprehensive image processing toolbox and numerous embedded mathematical functions facilitating statistical analysis.

a) Building b) Cycle c) Girl

d) Door e) Hats f) House

g) Parrots h) Window i) Plane

j) Boat k) Statue l) Lighthouse m) Woman

Figure 1: Images used in the experiment.

The comparison involved thirteen original and distorted images: building, cycle, girl, door, hats, house, parrots, window, plane, boat, statue, lighthouse, and woman, see Fig.1. These images were obtained from the "Lossless True Color Image Suite" provided by the "LIVE Image Quality Assessment Database" at the Laboratory of Image and Video Engineering, University of Texas, Austin regarding reference [12].

4. Results and Discussion

All used Image quality metrics are objective measurements that are automatics and mathematically defined algorithms. After applying some distortion (contrast enhancement) to the original nine images, see Fig.1, we got the distorted nine images included, and the image quality is applied to these distorted images and the results are compared. Measuring image quality for the nine images gave the results included in Table 1.

It can be seen from Table 1 in the 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. It is noticed that the result given by FSIM is closer to 1 than SSIM since it uses more features and involves better computations.

Table 1: Summary for different image quality metrics (MSE, PSNR, SSIM, FSIM).

Image	Gaussian Variance	MSE	PSNR	SSIM	FSIM
Boat	0.32	58.36	30.47	0.90	0.95
Building	0.61	351.95	22.67	0.77	0.95
Cycle	0.69	206.65	24.98	0.82	0.95
Door	0.34	45.51	31.55	0.85	0.96
Girl	0.33	64.23	30.05	0.89	0.93
Hats	0.30	40.03	32.11	0.91	0.95
House	0.42	76.75	29.28	0.84	0.97
Lighthouse	0.40	127.47	27.08	0.83	0.99
Parrots	0.28	37.02	32.45	0.94	0.97
Plane	0.29	81.07	29.04	0.90	0.95
Statue	0.37	53.23	30.87	0.91	0.96
Window	0.36	51.20	31.04	0.93	0.98
woman	0.54	131.70	26.93	0.83	0.96

5. 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 being time - consuming and expensive it still does better than objective IQM, and the objective IQM field is still open and needs lots of work to co - operate with subjective IQM.

6. Future Scope

We will work on defining other image quality assessment metrics trying to reach even better results in indicating image quality closer to human vision.

References

[1] Lanjiang Wang, A Survey on Image Quality Assessment, Cornell University, 2022, pp.1 - 14, https: //doi. org/10.48550/arXiv.2109.00347

[2] Domonkos Varga, A Combined Full - Reference Image Quality Assessment Method Based on Convolutional Activation Maps, MDPI algorithms, Hungary, 2020, 13 (12), https: //doi. org/10.3390/a13120313

[3] Mean Squared Error. wikipedia The Free Encyclopedia. [Online] [Cited: 29 2 2024.] https: //en. wikipedia. org/wiki/Mean_squared_error.

[4] Renuka Deshpande, Lata Ragha, and Satyendra Sharma, Video Quality Assessment through PSNR Estimation for Different Compression Standards. Indonesian Journal of Electrical Engineering and Computer Science, 2018, 11 (3), pp.918 – 924, https: //DOI. org/10.11591/ijeecs. v11. i3

[5] Domonkos Varga, Full - Reference Image Quality Assessment Based on Grünwald–Letnikov Derivative, Image Gradients, and Visual Saliency, Ronin Institute, 2022, Vol.11, No.4, Montclair, NJ 07043, USA.

[6] Domonkos Varga, A Human Visual System Inspired No - Reference Image Quality Assessment Method Based on Local Feature Descriptors, Ronin Institute, 2022, Vol.22, No.18, Montclair, NJ 07043, USA, https: //DOI. org/10.1117/12.862431

[7] Zhou Wang, Alan Bovik, A Universal Image Quality Index, Signal Processing Letters, 2002, Vol.9, pp.81 – 84, IEEE Xplore, https: //DOI. org/10.1109/97.995823

[8] Vicky Mudeng, Minseok Kim, and Se - woon Choe, Prospects of Structural Similarity Index for Medical Image Analysis, Signal, Applied Science, 2022, 12 (8), pp.1 - 34.

[9] Jacob Søgaard, Lukáš, Krasula Lukáš, and Muhammad, Shahid, Applicability of Existing Objective Metrics of Perceptual Quality for Adaptive Video Streaming. Conference: IS&T Electronic Imaging, Image Quality and System Performance XIII, 2016, San Francisco, https: //DOI. org/10.2352/ISSN.2470 - 1173.2016.13. IQSP - 206

[10] Umme Sara, Morium Akter, Mohammad Shorif Uddin, Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study, Journal of Computer and Communications, 2019, Vol.7, No.3, pp.8 - 19, https: //doi. org/10.4236/jcc.2019.73002

[11] Yusra Al Najjar, Effect of Contrast Measures on the Performance of No - Reference Image Quality Assessment Algorithm for Contrast - Distorted Images, Jordan Journal of Electrical Engineering, 2021, Vol.7, No.4, 390 – 404, https: //orcid. org/0000 - 0002 - 3369 - 4999

[12] TEXAS The University of Texas at Austin, Laboratory for Image & Video Engineering, [Online] [Cited: 1 3 2024.] https: //live. ece. utexas. edu/research/quality/subjective. htm,

Author Profile

Yusra Al Najjar received the B. S. from Kuwait University, and M. S. degrees in Computer Science from Al Balqua'a Applied University in 2008, respectively. And had the Ph. D. from University Tenaga National in Malysia. During 2009 - 2014, she worked as a lecturer in King Faisal University, Ministry of Education as a teacher and a supervisor, and from 2019 till now she is a lecturer at Taibah University in Madinah, Saudi Arabia.