# Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study

**Umme Sara[1], Morium Akter[2], Mohammad Shorif Uddin[2]** (ORCID)

[1]National Institute of Textile Engineering and Research, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh
Email: ummesara@niter.edu.bd, moriumcse@gmail.com, shorifuddin@juniv.edu

## Abstract

Quality is a very important parameter for all objects and their functionalities. In image-based object recognition, image quality is a prime criterion. For authentic image quality evaluation, ground truth is required. But in practice, it is very difficult to find the ground truth. Usually, image quality is being assessed by full reference metrics, like MSE (Mean Square Error) and PSNR (Peak Signal to Noise Ratio). In contrast to MSE and PSNR, recently, two more full reference metrics SSIM (Structured Similarity Indexing Method) and FSIM (Feature Similarity Indexing Method) are developed with a view to compare the structural and feature similarity measures between restored and original objects on the basis of perception. This paper is mainly stressed on comparing different image quality metrics to give a comprehensive view. Experimentation with these metrics using benchmark images is performed through denoising for different noise concentrations. All metrics have given consistent results. However, from representation perspective, SSIM and FSIM are normalized, but MSE and PSNR are not; and from semantic perspective, MSE and PSNR are giving only absolute error; on the other hand, SSIM and PSNR are giving perception and saliency-based error. So, SSIM and FSIM can be treated more understandable than the MSE and PSNR.

## Keywords

Image Quality, Computer Simulation, Gaussian Noise, Denoising

## 1. Introduction

Image Quality Assessment (IQA) is considered as a characteristic property of an image. Degradation of perceived images is measured by image quality assessment. Usually, degradation is calculated compared to an ideal image known as

reference image.

Quality of image can be described technically as well as objectively to indicate the deviation from the ideal or reference model. It also relates to the subjective perception or prediction of an image [1], such as an image of a human look.

The reduction of the quality of an image is affected by the noise. This noise depends on how it correlates with the information the viewer seeks in the image.

Visual information can have many featuring steps such as acquisition, enhancement, compression or transmission. Some information provided by the features of an image can be distorted after completion of the processing. That's why the quality should be evaluated by the human view perceptron [2]. Practically, there are two kinds of evaluation: subjective and objective. Subjective evaluation is time-consuming and also expensive to implement. Afterwards, the objective image quality metrics are developed on the basis of different aspects.

There are several techniques and metrics available to be used for objective image quality assessment. These techniques are grouped into two categories based on the availability of a reference image [3]. The categories are as follows:

1) Full-Reference (FR) approaches: The FR approaches focus on the assessment of the quality of a test image in comparison with a reference image. This reference image is considered as the perfect quality image that means the ground truth. For example, an original image is compared to the JPEG-compressed image [3] [4].

2) No-Reference (NR) approach: The NR metrics focus on the assessment of the quality of a test image only. No reference image is used in this method [3].

Image quality metrics are also categorized to measure a specific type of degradation such as blurring, blocking, ringing, or all possible distortions of signals.

The mean squared error (MSE) is the most widely used and also the simplest full reference metric which is calculated by the squared intensity differences of distorted and reference image pixels and averaging them with the peak signal-to-noise ratio (PSNR) of the related quantity [5].

Image quality assessment metrics such as MSE, PSNR are mostly applicable as they are simple to calculate, clear in physical meanings, and also convenient to implement mathematically in the optimization context. But they are sometimes very mismatched to perceive visual quality and also are not normalized in representation. With this view, researchers have taken-into account, two normalized reference methods to give structural and feature similarities. Structured similarity indexing method (SSIM) gives normalized mean value of structural similarity between the two images and feature similarity indexing method (FSIM) gives normalized mean value of feature similarity between the two images. All these are full-reference image quality measurement metrics.

In this paper, we compare the FSIM, SSIM, MSE and PSNR values between the two images (an original and a recovered image) from denoising for different noise concentrations.

## 2. Quality Measurement Technique

There are so many image quality techniques largely used to evaluate and assess the quality of images such as MSE (Mean Square Error), UIQI (Universal Image Quality Index), PSNR (Peak Signal to Noise Ratio), SSIM (Structured Similarity Index Method), HVS (Human Vision System), FSIM (Feature Similarity Index Method), etc. In this paper, we have worked on SSIM, FSIM, MSE and PSNR methods to find their suitability.

### 2.1. MSE (Mean Square Error)

MSE is the most common estimator of image quality measurement metric. It is a full reference metric and the values closer to zero are the better.

It is the second moment of the error. The variance of the estimator and its bias are both incorporated with mean squared error. The MSE is the variance of the estimator in case of unbiased estimator. It has the same units of measurement as the square of the quantity being calculated like as variance. The MSE introduces the Root-Mean-Square Error (RMSE) or Root-Mean-Square Deviation (RMSD) and often referred to as standard deviation of the variance.

The MSE can also be said the Mean Squared Deviation (MSD) of an estimator. Estimator is referred as the procedure for measuring an unobserved quantity of image. The MSE or MSD measures the average of the square of the errors. The error is the difference between the estimator and estimated outcome. It is a function of risk, considering the expected value of the squared error loss or quadratic loss.

Mean Squared Error (MSE) between two images such as $g(x, y)$ and $\hat{g}(x, y)$ is defined as [6]

$$\text{MSE} = \frac{1}{MN} \sum_{n=0}^{M} \sum_{m=1}^{N} \left[ \hat{g}(n, m) - g(n, m) \right]^2 \qquad (1)$$

From Equation (1), we can see that MSE is a representation of absolute error.

### 2.2. RMSE (Root Mean Square Error)

Root Mean square Error is another type of error measuring technique used very commonly to measure the differences between the predicted value by an estimator and the actual value. It evaluates the error magnitude. It is a perfect measure of accuracy which is used to perform the differences of forecasting errors from the different estimators for a definite variable [7].

Let us suppose that $\hat{\theta}$ be an estimator with respect to a given estimated parameter $\theta$, the Root Mean Square Error is actually the square root of the Mean Square Error as

$$\text{RMSE}(\hat{\theta}) = \sqrt{\text{MSE}(\hat{\theta})} \qquad (2)$$

### 2.3. PSNR (Peak Signal to Noise Ratio)

PSNR is used to calculate the ratio between the maximum possible signal power

and the power of the distorting noise which affects the quality of its representation. This ratio between two images is computed in decibel form. The PSNR is usually calculated as the logarithm term of decibel scale because of the signals having a very wide dynamic range. This dynamic range varies between the largest and the smallest possible values which are changeable by their quality.

The Peak signal-to-noise ratio is the most commonly used quality assessment technique to measure the quality of reconstruction of lossy image compression codecs. The signal is considered as the original data and the noise is the error yielded by the compression or distortion. The PSNR is the approximate estimation to human perception of reconstruction quality compared to the compression codecs.

In image and video compression quality degradation, the PSNR value varies from 30 to 50 dB for 8-bit data representation and from 60 to 80 dB for 16-bit data. In wireless transmission, accepted range of quality loss is approximately 20 - 25 dB [8].

PSNR is expressed as:

$$PSNR = 10\log_{10}\left(peakval^2\right)\!\big/MSE \tag{3}$$

Here, peakval (Peak Value) is the maximal in the image data. If it is an 8-bit unsigned integer data type, the peakval is 255 [8]. From Equation (3), we can see that it is a representation of absolute error in dB.

## 2.4. Structure Similarity Index Method (SSIM)

Structural Similarity Index Method is a perception based model. In this method, image degradation is considered as the change of perception in structural information. It also collaborates some other important perception based fact such as luminance masking, contrast masking, etc. The term structural information emphasizes about the strongly inter-dependant pixels or spatially closed pixels. These strongly inter-dependant pixels refer some more important information about the visual objects in image domain. Luminace masking is a term where the distortion part of an image is less visible in the edges of an image. On the other hand contrast masking is a term where distortions are also less visible in the texture of an image. SSIM estimates the perceived quality of images and videos. It measures the similarity between two images: the original and the recovered.

There is an advanced version of SSIM called Multi Scale Structural Similarity Index Method (MS-SSIM) that evaluates various structural similarity images at different image scale [9]. In MS-SSIM, two images are compared to the scale of same size and resolutions. As Like as SSIM, change in luminance, contrast and structure are considered to calculate multi scale structural similarity between two images [10]. Sometimes it gives better performance over SSIM on different subjective image and video databases.

Another version of SSIM, called a three-component SSIM (3-SSIM) that corresponds to the fact: human visual system observes the differences more accurately in textured regions than the smooth regions. This 3-component SSIM

model was proposed by Ran and Farvardin [11] where an image is disintegrated into three important properties such as edge, texture and smooth region. The resulting metric is calculated as a weighted average of structural similarity for these three categories. The proposed weight measuring estimations are 0.5 for edges, 0.25 for texture and 0.25 for smooth regions. It can also be mentioned that a 1/0/0 weight measurement influences the results to be closer to the subjective ratings. This can be implied that, no textures or smooth regions rather edge regions play a dominant role in perception of image quality [11].

## 2.5. DSSIM (Structural Dissimilarity)

There is another distance metric referred as Structural Dissimilarity (DSSIM) deduced from the Structural Similarity (SSIM) can be expressed as:

$$\text{DSSIM}(x,y) = \frac{1 - \text{SSIM}(x,y)}{2} \tag{4}$$

The SSIM index method, a quality measurement metric is calculated based on the computation of three major aspects termed as luminance, contrast and structural or correlation term. This index is a combination of multiplication of these three aspects [12].

Structural Similarity Index Method can be expressed through these three terms as:

$$\text{SSIM}(x,y) = \left[ l(x,y) \right]^{\alpha} \cdot \left[ c(x,y) \right]^{\beta} \cdot \left[ s(x,y) \right]^{\gamma} \tag{5}$$

Here, $l$ is the luminance (used to compare the brightness between two images), $c$ is the contrast (used to differ the ranges between the brightest and darkest region of two images) and $s$ is the structure (used to compare the local luminance pattern between two images to find the similarity and dissimilarity of the images) and $\alpha$, $\beta$ and $\gamma$ are the positive constants [13].

Again luminance, contrast and structure of an image can be expressed separately as:

$$l(x,y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{6}$$

$$c(x,y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{7}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \tag{8}$$

where $\mu_x$ and $\mu_y$ are the local means, $\sigma_x$ and $\sigma_y$ are the standard deviations and $\sigma_{xy}$ is the cross-covariance for images $x$ *and* $y$ *sequentially*. If $\alpha = \beta = \gamma = 1$, then the index is simplified as the following form using Equations (6)-(8):

$$\text{SSIM}(x,y) = \frac{\left(2\mu_x \mu_y + C_1\right)\left(2\sigma_x \sigma_y + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{9}$$

From Equation (9) we can see that FSIM is in normalized scale (values between 0 to 1). We can also express it in dB scale as $10\log_{10}[\text{SSIM}(x, y)]$.

## 2.6. Features Similarity Index Matrix (FSIM)

Feature Similarity Index Method maps the features and measures the similarities between two images. To describe FSIM we need to describe two criteria more clearly. They are: Phase Congruency (PC) and Gradient Magnitude (GM).

Phase Congruency (PC): A new method for detecting image features is phase congruency. One of the important characteristics of phase congruency is that it is invariant to light variation in images. Besides, it is also able to detect more some interesting features. It stresses on the features of the image in the domain frequency. Phase congruency is invariant to contrast.

Gradient magnitude (GM): The computation of image gradient is a very traditional topic in the digital image processing. Convolution masks used to express the operators of the gradient. There are many convolutional masks to measure the gradients. If $f(x)$ is an image and $G_x$, $G_y$ of its horizontal and vertical gradients, respectively. Then the gradient magnitude of $f(x)$ can be defined as [13]

$$\sqrt{G_x^2 + G_y^2} \tag{10}$$

In this paper we are going to calculate the similarity between two images to assess the quality of images. Let two images are $f_1$ (test image) and $f_2$ (reference image) and their phase congruency can be denoted by $PC_1$ and $PC_2$, respectively. The Phase Congruency (PC) maps extracted from two images $f_1$ and $f_2$ and the Magnitude Gradient (GM) maps $G_1$ and $G_2$ extracted from the two images too. FSIM can be defined and calculated based on $PC_1$, $PC_2$, $G_1$ and $G_2$. At first, we can calculate the similarity of these two images as

$$S_{PC} = \frac{2PC_1 PC_2 + T_1}{PC_1^2 + PC_2^2 + T_1} \tag{11}$$

where $T_1$ is a positive constant which increases the stability of $S_{pc}$. Practically $T_1$ can be calculated based on the PC values. The above equation describes the measurement to determine the similarity of two positive real numbers and its range is within 0 to 1.

Similarly, we can calculate the similarity from $G_1$ and $G_2$ as

$$S_G = \frac{2G_1 G_2 + T_2}{G_1^2 + G_2^2 + T_2} \tag{12}$$

where $T_2$ is a positive constant which depends on the dynamic range of gradient magnitude values. In this paper, both $T_1$ and $T_2$ are constant so that the FSIM can be conveniently used.

Now $S_{PC}$ and $S_G$ are combined together to calculate the similarity $S_L$ of $f_1$ and $f_2$. $S_l$ can be defined as

$$S_L(x) = \left[S_{PC}(x)\right]^{\alpha} \cdot \left[S_G(x)\right]^{\beta} \tag{13}$$

where the parameters $\alpha$ and $\beta$ are used to adjust the relative importance of PC

and GM features. In this paper, we set $\alpha = \beta = 1$ for convenience. From Equations (11) to (13), it is evident that FSIM is normalized (values between 0 to 1).

## 3. Experimental Results and Discussions

We used three benchmark images (Lena, Barbara, Cameraman) and then used Gaussian noise of different concentrations (noise variances). For all noise variances the mean is taken as 1.0. Then we applied Gaussian convolutional mask for denoising. The original, noisy and denoised images are shown in Figure 1, Figure 2 and Figure 3 for noise variances 0.2, 0.4 and 0.6, respectively.



**Figure 1.** ((a), (d), (g)) are three bench mark original images; ((b), (e), (h)) are the corresponding noisy images with noise variance 0.2; ((c), (f), (i)) are the corresponding denoised images. Here, the noise level is 0.4. The MSE, PSNR, SSIM and FSIM value of two images (the original and the recovered image) are listed in Table 1.
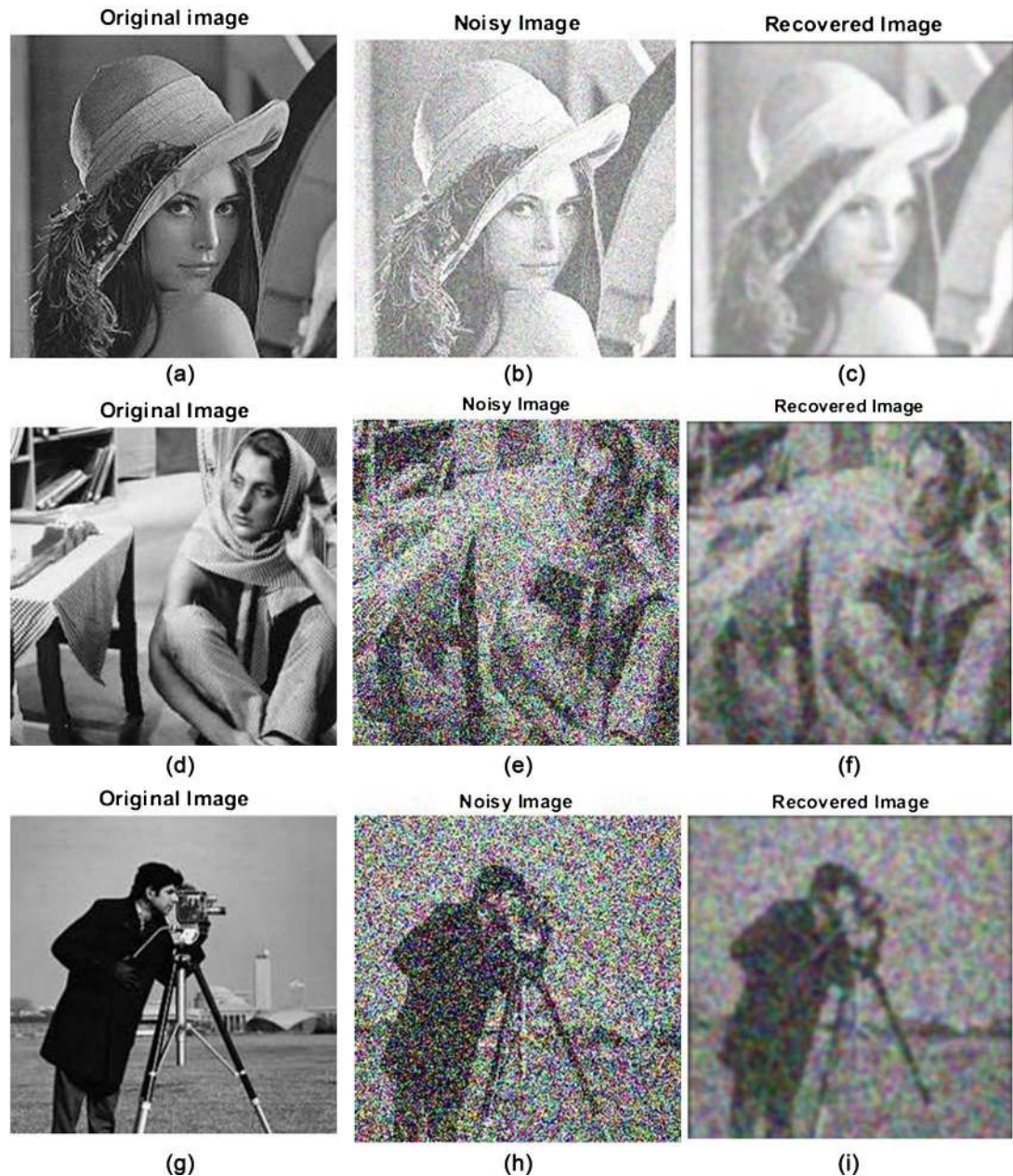
**Figure 2.** ((a), (d), (g)) are three bench mark original images; ((b), (e), (h)) are the corresponding noisy images with noise variance 0.4; ((c), (f), (i)) are the corresponding denoised images Here, the noise level is 0.4. The MSE, PSNR, SSIM and FSIM value of two images (the original and the recovered image) are listed in Table 1.

After denoising, we estimated the quality of the denoised (restored/recoverd) images by using FSIM, SSIM, PSNR and MSE metrics. The summary of quality matrices calculation is shown in Table 1. From this table, we can see that all metrics have given almost consistent results. However, from representation perspective, SSIM and FSIM are normalized, but MSE and PSNR are not. So, SSIM and FSIM can be treated more understandable than the MSE and PSNR. This is because, MSE and PSNR are absolute errors, however, SSIM and FSIM are giving perception and saliency-based errors. If noise level is increasing, then the

**Figure 3.** ((a), (d), (g)) are three bench mark original images; ((b), (e), (h)) are the corresponding noisy images with noise variance 0.6; ((c), (f), (i)) are the corresponding denoised images Here, the noise level is 0.6. The MSE, PSNR, SSIM and FSIM value of two images (the original and the recovered image) are listed in Table 1.

recovery quality of output image is also deteriorating.

## 4. Conclusion

Image Quality Assessment plays a very significant role in digital image processing applications. The metrics (MSE, PSNR, SSIM and FSIM) are applied in this paper to get the best quality metric. We have done simulating experiments using Gaussian noise through Gaussian filtering technique. The obtained image quality has been judged on applying the above metrics. We found consistent

**Table 1.** Error deduction summary for different image quality metrics (MSE, PSNR, SSIM, FSIM).

| Image | Noise Level (Gaussian-variance) | Quality Assessment Techniques | | | |
|---|---|---|---|---|---|
| | | MSE | PSNR | SSIM | FSIM |
| | 0.2 | 21.56 | 21.54 | 0.78 | 0.89 |
| Lena | 0.4 | 16.81 | 16.81 | 0.74 | 0.86 |
| | 0.6 | 14.18 | 14.18 | 0.7 | 0.8 |
| | 0.2 | 21.95 | 21.95 | 0.8 | 0.88 |
| Barbara | 0.4 | 17.8 | 17.79 | 0.75 | 0.84 |
| | 0.6 | 15.49 | 15.48 | 0.7 | 0.8 |
| | 0.2 | 21.64 | 21.65 | 0.73 | 0.84 |
| Cameramen | 0.4 | 17.29 | 17.31 | 0.76 | 0.86 |
| | 0.6 | 15.27 | 15.28 | 0.76 | 0.87 |

results for all the metrics. However, from representation perspective, SSIM and FSIM are normalized, but MSE and PSNR are not. So, SSIM and FSIM can be treated more understandable than the MSE and PSNR. This is due to the fact that MSE and PSNR are absolute errors, however, SSIM and FSIM are giving perception and saliency-based errors. If noise level is increasing, then the recovery quality of output image is also deteriorating. So, we can conclude that SSIM and SSIM are comparatively better than MSE and PSNR metrics from human visual perspective.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Thung, K.-H. and Raveendran, P. (2009) A Survey of Image Quality Measures. *IEEE Technical Postgraduates* (*TECHPOS*) *International Conference*, Kuala Lumpur, 14-15 December 2009, 1-4.

[2] Oszust, M. (2016) Full-Reference Image Quality Assessment with Linear Combination of Genetically Selected Quality Measures. Yongtang Shi, Nankai University, China, Vol. 11, No. 6, June 24.

[3] Wang, Z. and Simoncelli, E.P. (2005) Reduced-Reference Image Quality Assessment Using A Wavelet-Domain Natural Image Statistic Model. *Human Vision and Electronic Imaging X, Proc. SPIE*, Vol. 5666, San Jose, CA, 18 March 2005. https://doi.org/10.1117/12.597306

[4] Lahoulou, A., Larabi, M.C., Beghdadi, A., Viennet, E. and Bouridane, A. (2016) Knowledge-Based Taxonomic Scheme for Full-Reference Objective Image Quality Measurement Models. *Journal of Imaging Science and Technology*, **60**, 1-15.

[5] Wang, Z. and Sheikh, H.R. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, No. 4. https://doi.org/10.1109/TIP.2003.819861

[6] Søgaard, J., Krasula, L., Shahid, M., Temel, D., Brunnstrom, K. and Razaak, M. (2016) Applicability of Existing Objective Metrics of Perceptual Quality for Adaptive Video Streaming. *Society for Imaging Science and Technology IS&T International Symposium on Electronic Imaging.*

[7] Mean Squared Error. https://math.tutorvista.com/statistics/mean-squared-error.html

[8] Deshpande, R.G., Ragha, L.L. and Sharma, S.K. (2018) Video Quality Assessment through PSNR Estimation for Different Compression Standards. *Indonesian Journal of Electrical Engineering and Computer Science*, **11**, 918-924.

[9] Wang, Z., Simoncelli, E.P. and Bovik, A.C. (2004) Multiscale Structural Similarity for Image Quality Assessment. *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, **2**, 1398-1402.

[10] Dosselmann, R. and Yang, X.D. (2011) A Comprehensive Assessment of the Structural Similarity Index. *Signal, Image and Video Processing*, **5**, 81-91. https://doi.org/10.1007/s11760-009-0144-1

[11] Li, C.F. and Bovik, A.C. (2009) Three-Component Weighted Structural Similarity Index. *Image Quality and System Performance VI, SPIE Proc.* 7242, San Jose, CA, 19 January 2009, 1-9.

[12] Brooks, A.C., *et al.* (2008) Structural Similarity Quality Metrics in a Coding Context: Exploring the Space of Realistic Distortions. *IEEE Transactions on Image Processing*, **17**, 1261-1273. https://doi.org/10.1109/TIP.2008.926161

[13] Kumar, R. and Moyal, V. (2013) Visual Image Quality Assessment Technique Using FSIM. *International Journal of Computer Applications Technology and Research*, **2**, 250-254. https://doi.org/10.7753/IJCATR0203.1008