APPLICATIONS OF
MODELLING AND SIMULATION
http://arqiipubl.com/ams eISSN 2600-8084 VOL 6, 2022, 10-19

Improving Measurement Bias of Structural Similarity Index (SSIM) using Absolute Difference Equation

Muhammad Irfan Jaafar*, Sophan Wahyudi Nawawi and Ruzairi Abdul Rahim

School of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

*Corresponding author: muhammadirfanjaafar96@gmail.com

Submitted 22 November 2021, Revised 02 January 2022, Accepted 10 January 2022, Available online 19 January 2022.

Abstract: Structural similarity index (SSIM) is a framework for assessing the perceptual quality from an image using the degrading and structural information of an image. It is an alternative method for quantifying visual image quality since subjective evaluation by humans, in practice, is too inconvenient, time-consuming and expensive. SSIM has been widely used for almost two decades in different research disciplines in quality assessment. However, there is a deficiency in some of the mathematical components in the SSIM that may lead to incorrect, impractical and unrealistic results. In this paper, we address the problems with the SSIM and propose replacing the luminance and contrast component with absolute difference equations that obeys Weber's law of human perception of change. We compare the results of both SSIM and new proposed SSIM in the luminance, contrast and human perception test. The proposed SSIM seem to be more human-like in determining image quality compared to the previous version of SSIM.

Keywords: Contrast; Luminance; Structural similarity index (SSIM); Structure; Weber-Fechner law.

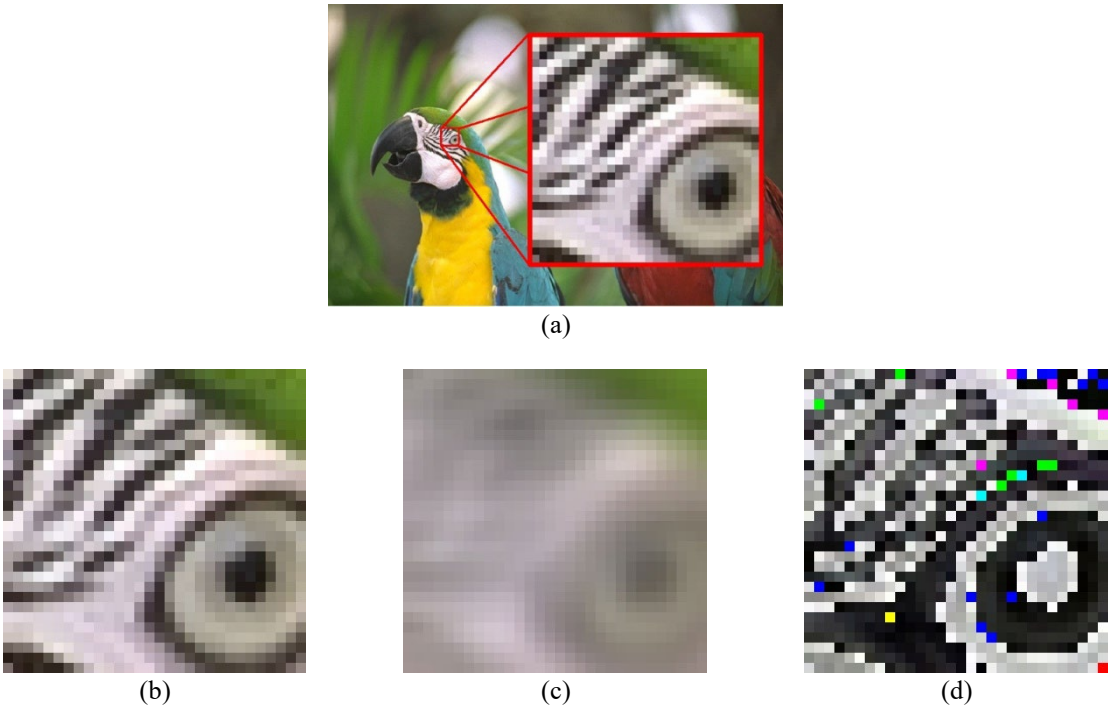
1. INTRODUCTION

The SSIM [1] was published in April 2004 and has more than 30,000 citations in Google Scholar. SSIM has been a popular choice of technique when comparing the similarities between the two images. It perceives changes in structural information of the image unlike most of the predecessor methods such as Mean Square Error (MSE) and Peak signal-to-noise ratio (PSNR) computing perceived errors. It is also a lot closer to human perception of quality compared to others. Since then, many forms and variances of SSIM have been introduced, like Multi-Scale SSIM [2], Multi-component SSIM [3], Complex Wavelet SSIM [4] and many more [5, 6, 7].

Before SSIM, MSE was and is still being used to measure the similarity or level of error between two signals or image processing, because of how easy and inexpensive it is to compute. But absolute errors techniques are very sensitive to distorted, shifted, rotated or zoomed images. The value of MSE relative to the original is huge even though the image looks almost identical to human eyes after distortion. SSIM is based on human perceptual visual stimulation. The human visual system is an expert in extracting structural information from the visual scene. The main idea of SSIM is to measure the similarity of luminance, contrast and structure from the local patch. All these local similarities are then being combined to form local SSIM and the total patch of local SSIM is called SSIM map where it can be used to localize the error in images. SSIM is proven to be more robust and human-like in quality assessment compared to MSE.

2. MATHEMATICAL COMPONENT

SSIM compares local patterns of pixel intensities that have been normalized for luminance, contrast and structural. It is based on the statistical measurements such as mean μ, standard deviation σ, and covariance σ_{xy}. Two non-negative image signals, x and y, have been aligned spatial patches extracted from each image. Mean measurement on an image can be done using the sliding window method by replacing the centre value of the window with the average pixel value in the local window. The window can be of any size or shape depending on the user. It is also worth mentioning that mean measurement is not limited to the averaging pixels value only, it can be median, Gaussian or any method that can remove other details from images except for luminance information. Standard deviation measurement on an image is the process of finding the absolute difference between the original image with the mean measurement for every pixel value. Figure 1 illustrates the results of mean and standard deviation measurements on the local patch. Covariance measurement involves the multiplication of the results of the difference between the original image with its mean measurement from two images.



(a) (b) (c) (d)

Figure 1. Visualization of image signal under mean and standard deviation measurement. (a) A sample of a patch taken from an image, (b) Original patch, (c) Patch of mean, μ, (d) Patch of standard deviation, σ.

These patches are then being compared for luminance and contras. The luminance equation between 2 patches is given by:

$$l(x,y) = \frac{(2\mu_x\mu_y + C_1)}{({\mu_x}^2 + {\mu_y}^2 + C_1)} \tag{1}$$

where μ_x and μ_y are the means intensity of the image x and y and constant C_1 is included to avoid instability when ${\mu_x}^2 + {\mu_y}^2$ close to zero. The second component is the contras equation, given by:

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

σ_x and σ_y are the standard deviations of intensity of the image x and y and σ_{xy} is the covariance of the image x and y. Similarly, C_2 is included to avoid instability when ${\sigma_x}^2 + {\sigma_y}^2$ is close to zero. The structural equation is given by:

$$s(x,y) = \frac{(\sigma_{xy} + C_3)}{(\sigma_x\sigma_y + C_3)} \tag{3}$$

SSIM is the product of $l(x,y)$, $c(x,y)$ and $s(x,y)$, with the weighted function of α, β and γ, which result in:

$$SSIM(x,y) = \left[l(x,y)^\alpha \cdot c(x,y)^\beta \cdot s(x,y)^\gamma\right] \tag{4}$$

The original paper proposes that $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$ where L is 255 for 8-bit images and 1 for 2-bit images and constant C_3 is equal to $C_2/2$. K_1 and K_2 were chosen as 0.01 and 0.03 respectively. With this, C_1 equal to 6.5025 and C_2 equal to 58.5225 for 8-bit images. A simplified equation of SSIM when the weights of α, β and γ are set to 1 is given by:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{({\mu_x}^2 + {\mu_y}^2 + C_1)({\sigma_x}^2 + {\sigma_y}^2 + C_2)} \tag{5}$$

There are several problems with the usage of formula in the first and the second components, l and c. These formulas can be represented in a form of:

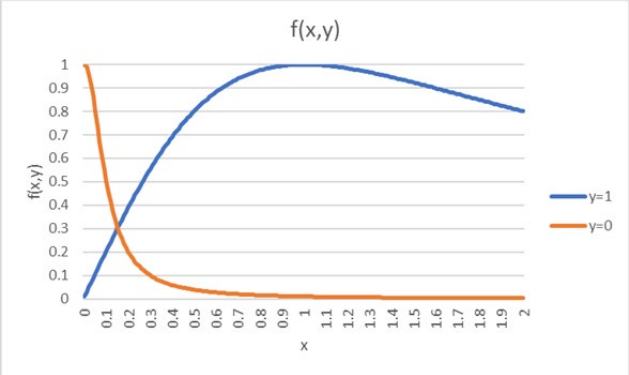


Figure 2. Results of $f(x, y)$ for x in range of 0 to 2

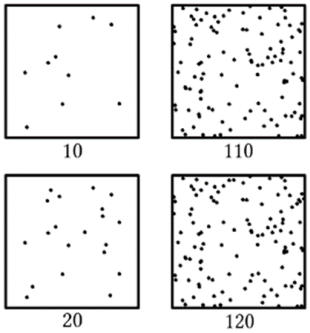


Figure 3. Illustration of the Weber-Fechner law

$$f(x, y) = \frac{2xy + k}{x^2 + y^2 + k} \tag{6}$$

The equation used will only return 1 if both variables x and y are similar and become approximately 0 if either one of the variables is 0 and k is small. The closer both values are together, the closer the result is to 1 and further apart both values, the closer the result to 0. The similarity is measured by the function of $f(x, y)$ were created to fulfil all the following conditions:

1. Symmetry: $f(x, y) = f(y, x)$.
2. Boundedness: $f(x, y) \leq 1$.
   Unique maximum: $f(x, y) = 1$ if $x = y$ (for all positive numbers)

Figure 2 shows the result of $f(x, y)$ where x is a value from 0 to 2 and y is 1 (in Blue) and y is 0 (in Orange) with a constant k of 0.01. The result of the equation shows that not only it is nonlinear, but it also has bias result where the result (in Blue) when comparing between high values (x between 1 to 2) are closer to 1 than the result comparing between low values (x between 0 to 1). The asymmetrical graph is a confirmation where the mathematical relation used in the SSIM does not represent the spatial distance between the two variables, and to be fair, the authors of the SSIM never make any claim that it is a linear metric or has a spatial distance function.

The biggest problem with the equation $f(x, y)$ is that it does not perfectly represent Weber's law of human perception of change. Weber's law [8] stated that the ratio of the increment threshold to background intensity is constant. In other words, human perception to change is logarithmic to reference stimulus and the rate of noticeable difference ΔI over reference stimulus I is roughly constant.

$$k = \frac{\Delta I}{I} \tag{7}$$

Take Figure 3 as an example, when comparing the right images it is hard for humans to tell the difference since 110 dots and 120 dots are almost visually similar. On the left side, the amount of difference between 10 dots and 20 dots is a lot more noticeable. When applying these dots into the equation $f(x, y)$, the result will be 0.8 on the left side and 0.996 on the right side with constant k equal to 0. It is found that the result is 0.8, even though the difference is noticeable. It does satisfy qualitatively, but the equation $f(x, y)$ is too extreme. The usage of $x^2 + y^2$ as a stimulus or reference is pretty odd for humans brain to process. Simply put, formula $f(x, y)$ is not how humans comprehend similarity.

2.1 Luminance

Luminance is the luminous intensity of light emitted and it has its measurement unit in the scientific community and can be measured, but humans perceive it as brightness. In SSIM, it is a component to compare the local intensity or brightness between two images and value it in the range of 0 to 1. To measure the brightness of a scene, the local average brightness of two images can be used to represent the luminance of that particular area. Similar to signal processing, the mean of a signal is used instead of the actual signal because the actual signal may contain noise that will mess up the reading of the amplitude at a certain position.

The proposed solution for the luminance equation is to use the relative difference of the local mean between two images. Using the absolute difference between two mean intensities against the maximum range of intensity that humans can perceive which is the difference between total black to total white. Using luminance difference as noticeable difference ΔI and range of black to white as the initial stimulus I. The largest range of pixel intensity in digital grayscale images would be $2^n - 1$, where n is 8 for an 8-bit image. Since the relative difference is an error measurement, the result of it would be negatively correlated to the human perception, so 1 need to be subtracted to the relative difference to inverse it. The equation would be:

$$l(x,y) = 1 - \frac{|\mu_x - \mu_y|}{(2^n - 1)} \tag{8}$$

Constant may not be needed since $(2^n - 1)$ result in a large number, but it can be added if the user of the equation needs their relative different result to be thresholded or shifted.

2.2 Contrast

The contrast in the visual system is the difference between the light and dark areas of one image. It is more difficult to measure than luminance because it is about the scale of change or gradient in an image and different people may have different definitions for contrast. A common term used in describing a high contras image such as “pop” or “sharp” while “faded” or “washed out” is used to describe low contras. SSIM used local standard deviation as the contrast of the image because it shows the range of difference between intensity locally. This is true and aligns with the contrast commonly used for digital images which are the Root mean square (RMS) contrast where it takes the standard deviation of the pixel intensity.

Intensity can be measured by finding the absolute difference with the maximum intensity range as a stimulus because humans look at colour black and white in their everyday life and know how to perceive what is light and dark. But when comparing contrast, absolute difference against the maximum contrast range may not work, because how often do human see maximum contrast in their daily life, and do they know what maximum contrast look like. But when being shown two images, humans can spot the difference what is high and low contrast. So, the proposed contrast equation would be only the scale between low standard deviation with high standard deviation. Making the absolute difference between contrast as the amount of change ΔI and the maximum contrast between the two as stimulus I.

$$c(x,y) = \begin{cases} 1 - \dfrac{|\sigma_x - \sigma_y|}{\sigma_y + k}, & \sigma_x < \sigma_y \\ 1 - \dfrac{|\sigma_x - \sigma_y|}{\sigma_x + k}, & \sigma_x \geq \sigma_y \end{cases} \tag{9}$$

Equation (9) can be further simplified as:

$$c(x,y) = \begin{cases} \left|\dfrac{\sigma_x + k}{\sigma_y + k}\right|, & \sigma_x < \sigma_y \\ \left|\dfrac{\sigma_y + k}{\sigma_x + k}\right|, & \sigma_x \geq \sigma_y \end{cases} \tag{10}$$

Constant k is a small value to represent a baseline level of a reasonable standard deviation to be considered as local contrast in comparing two images. Based on some testing, k needs to be at least 1 to yield a good result since the standard deviation of x and y can be smaller than 1 and cause instability.

2.3 Structure

The structural component of SSIM is a derivation from the well-known Pearson’s correlation coefficient equation [9]. It is a very good tool to measure the relation between two variances either the two images have a positive relation, negative relation, or no relation at all. Structure in SSIM is also good at finding errors of texture in the images. For example, it can determine whether the image is pixelated compared to the original. Since there can be a negative correlation, the structure is the only part of the equations where it can return negative values, making SSIM has a possible range of -1 to 1. But this negativity can only happen when more than half of the image compared is negatively correlated (negative image).

Pearson’s correlation coefficient is stated to be not robust and sensitive to noise, the result can be misleading if any outliers value exists in the data. Constant in the structural formula is added to prevent instability when $\sigma_x\sigma_y$ is too small, both denominator and denominator are shifted by C_3, this constant subsequently act as the noise and negativity reduction. This paper used the same structure equation as the original SSIM proposed since the equation are capable of comparing the texture between images, and does not need any modification.

2.4 Proposed SSIM

The proposed SSIM is to remove Equation (6) from the luminance and contrast component of SSIM and replace it with something more linear, like absolute errors Equation (8) for luminance and a simple relative difference in Equation (10) for contrast. The proposed solution is also a multiplication of $l(x,y)$, $c(x,y)$ and $s(x,y)$, with the weighted function of α, β and γ, similar to Equation (4).

3. EXPERIMENTS ON SSIM

This paper uses the term SSIM referring to the original SSIM and New SSIM referring to the proposed SSIM. There are 3 parts of experiment which are luminance test, contrast test and human perception test comparing the result of SSIM with New

SSIM. Note that MSSIM is the mean result of the SSIM maps, which is also called the quality index in the experiment. All tests were done on Python environment with scikit-image's structural similarity as the SSIM for this experiment and the images used are 8-bit images.

Weights of α, β and γ are set to 1 and use 11×11 circular-symmetric Gaussian weighting function, $w = \{w_i | i = 1,2 \dots, N\}$, with a standard deviation of 1.5 for both SSIM. The local mean, standard deviation and covariance are as follows:

$$\mu_x = \sum_{i=1}^{N} w_i x_i \tag{11}$$

$$\sigma_x = \left(\sum_{i=1}^{N} w_i (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \tag{12}$$

$$\sigma_{xy} = \sum_{i=1}^{N} w_i (x_i - \mu_x)(y_i - \mu_y) \tag{13}$$

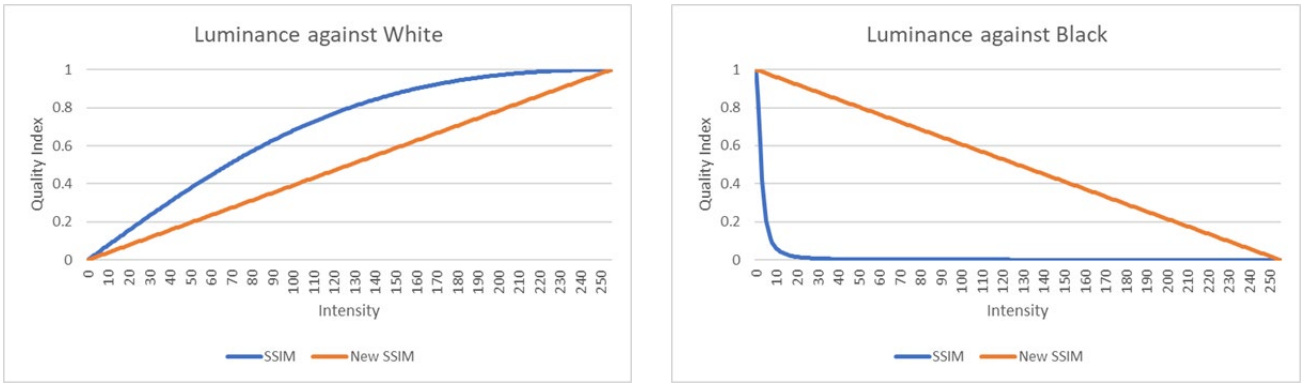
For SSIM, the constants are $K_1 = 0.01$ and $K_2 = 0.03$, ($C_1 = 6.5025$, $C_2 = 58.5225$ and $C_3 = 29.26125$). On the other hand, for the new SSIM, the constants are $k = 1$ and $C_3 = 29.26125$. In the human perception test, results from Complex wavelet structural similarity (CW-SSIM) and Multiscale structural similarity (MS-SSIM) are also presented in these comparisons. The source code used for the model of CW-SSIM and MS-SSIM can be found in [10, 11] with the default setting.

3.1 Luminance Test

The purpose of this test is to see how both SSIM respond when comparing two images with different intensities. Figure 4(a) is the results of MSSIM when a blank white image was compared to a grayscale image at all levels of intensity from 0 (black) to 255 (white). Each image has the same size of 300×300 pixels. In contrast, Figure 4(b) shows the results when comparing all levels of grayscale image to a blank black image using the same experiment. The result of the quality index for SSIM is non-linear, not only that, the curvature at all levels of intensity against brighter image is a lot 'soft curved' and anything above 160 of intensity against white easily score above 0.9. Contrary to that, SSIM is too responsive when compare with black because all intensity value above 10 is approximate 0.

Figure 5 shows results of a similar experiment [12] that has been done previously which stated the same problem, the intensity of 222 compared to intensity 255 wield a 0.99047 even though the brightness difference between the images is obviously visible, the result of ≥ 0.99 indicates that it is indistinguishable by the model. Yet when the intensity of 26 compared to intensity 0, the mean SSIM only score 0.00953, saying that the two images are very different even when the two images look almost similar. Results using new SSIM when comparing 222 with 255 is 0.87059, and the score for intensity 0 with 26 is 0.89804 because new SSIM measures relative error between the two intensities.

The result of the new SSIM behave as expected, the quality index increases linearly with intensity when compared to white and decreases linearly when compared to black. No bias score toward either bright image or dark image and the results imply that the new SSIM's intensity formula shows the spatial distance between the intensity. This is because that the new intensity equation is based on the mathematical relation of relative error.



(a) (b)

Figure 4. Comparing two images with different intensities. (a) Comparing image with intensity 0 to 255 with white image. (b) Comparing image with intensity 0 to 255 with black image

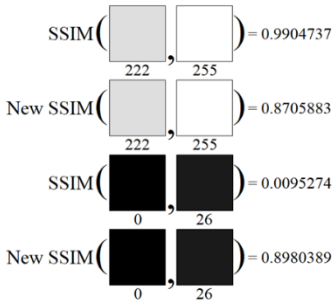
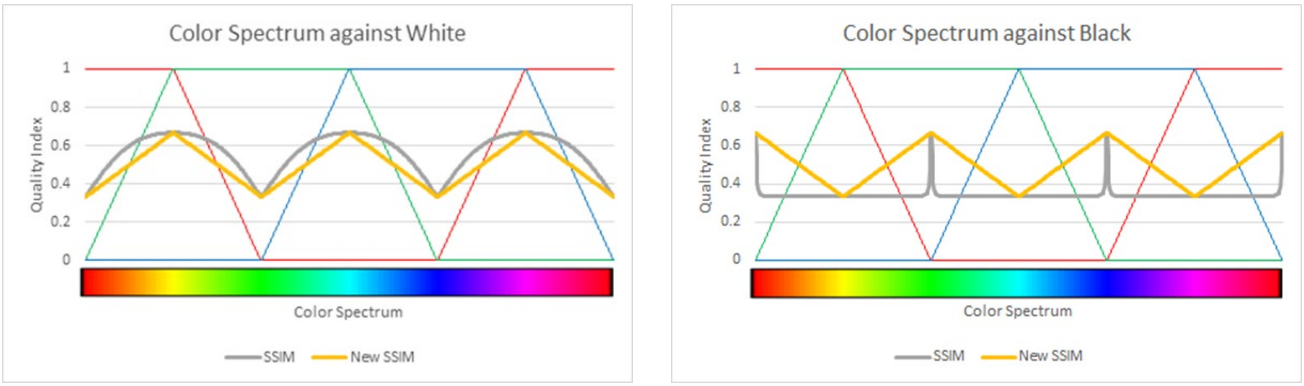


Figure 5. The mean result of SSIM and New SSIM



(a) (b)

Figure 6. Comparing two images with different colours. (a) Comparing images along the colour spectrum against white. (b) Comparing images along the colour spectrum against black

The original authors did a quality test on compressed colour images, but the colour component of the images does not significantly change so it should not change the performance of SSIM. The authors of [12] did a colour test by converting the colour image to grayscale before applying SSIM to it, a blank yellow colour (255 red, 255 green, 0 blue) give a score approximate to 0.99 when compared to white. A reason for this is that when converting yellow to grayscale, the intensity value becomes 170, and 170 will easily score above 0.9. It was advised to use a metric that handles colour inherently. Another approach described in the paper is to apply SSIM on YC_rC_b components individually.

In this paper, SSIM is applied to the RGB component separately. Then the average results of all three components were taken as shown in Figure 6. Figure 6(a) is the blank image with different colours against white while Figure 6(b) is the image with different colours against black. Only one colour component changes at a time. Based on the results, the yellow colour does not give a score approximate to 0.9 but instead, a score of ≈0.667 (both SSIM and new SSIM) indicate that the colour component does need to be handled separately rather than converted into grayscale first before comparing. Again, the same pattern when using SSIM can be seen, where it has a soft curve when compared to white while most of the scores stay flat (≈0.333) when compared to black, except for a certain colour (colours where 1 component is maximum intensity and 2 components minimum intensity).

The results of SSIM and new SSIM in the colour spectrum experiment shows a repeating pattern with symmetry shape in the x-axis of the graph proving that both models obey the symmetry properties. The graph pattern of the new SSIM for the colour spectrum against black is an inverse pattern for the white experiment since black and white is always on the opposite side of the luminance chart but the results of SSIM for colour spectrum against black does not reflect the results of white experiment further proving that SSIM defies the distance function. On other hand, the new SSIM's score changes linearly as the colour component changes linearly for both experiments, showing that it has the digital colour-spatial like feature.

3.2 Contrast Test

The contrast test is to see how sensitive SSIM and new SSIM when comparing images with different contrast. The test is done by comparing pixel-level black and white checkerboards with different contrast factor checkerboards. The contrast factor in this experiment used Python Imaging Library (PIL) where the enhancement factor of 0.0 gives a plain grey image and the factor of 1.0 gives the original image. The image size for this experiment is 300×300. The reason checkerboard image is used for contrast test is that it has the most optimum level of gradient, every pixel is different in terms of intensity compared to its nearest neighbours.

Figure 7 shows the pixel level checkerboard zoomed into 4×4 pixels for a different contrast. Comparisons of pixel checkerboard of different contrast with pixel checkerboard of contrast factor of 1 are shown in Figure 8. The SSIM increases rapidly at first, for low contrast images but slowly increases for higher contrast, while the new SSIM steadily increases as the contrast factor increases. New SSIM can be seen that is in line with the machine definition of contrast. Another test is done on the colour images, comparing the original image with the same image at different contrast factors.

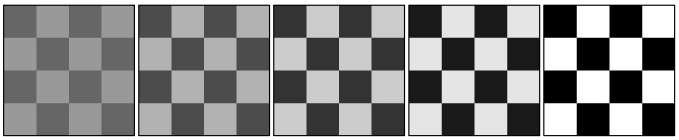


Figure 7. Zoom in 4x4 pixel checkerboard. From left to right, contrast factors are 0.2, 0.4, 0.6, 0.8 and 1.0

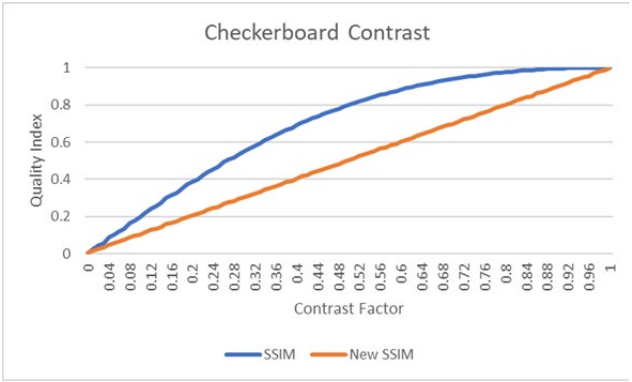
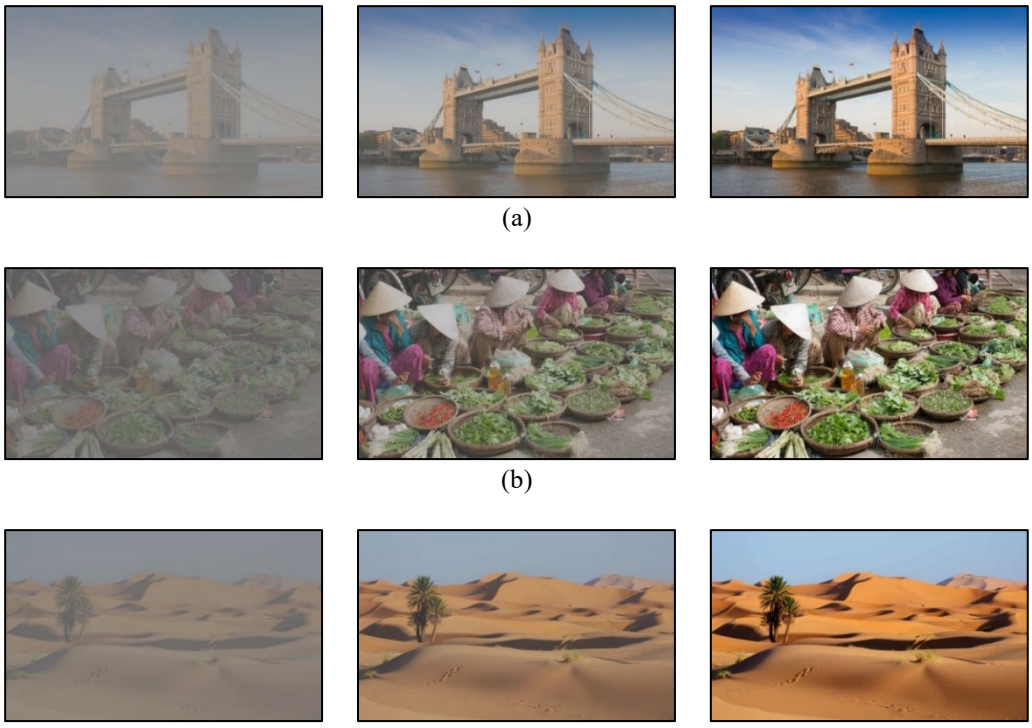


Figure 8. Comparing pixel checkerboard of different contrast with pixel checkerboard of contrast factor of 1

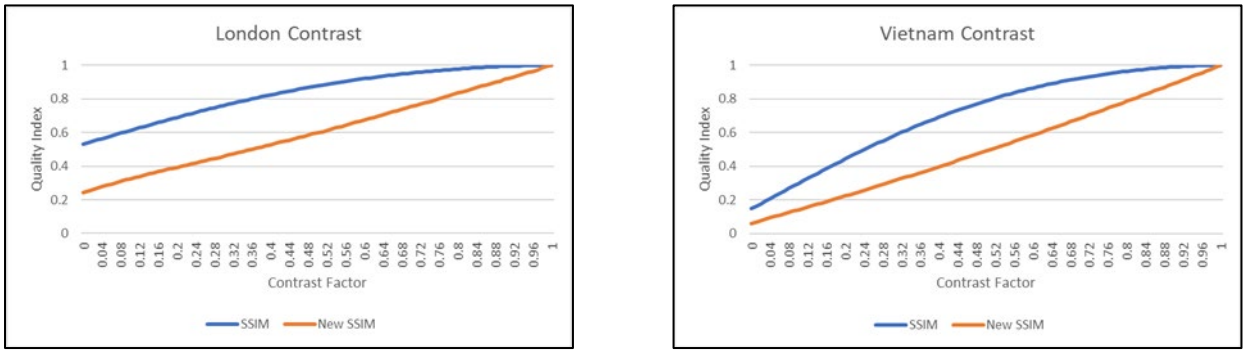


(a) (b) (c)

Figure 9. From left to right, images with different contrast factors of 0.2, 0.6 and 1.0. (a) London Bridge, (b) Vietnam Street, (c) Sahara Desert

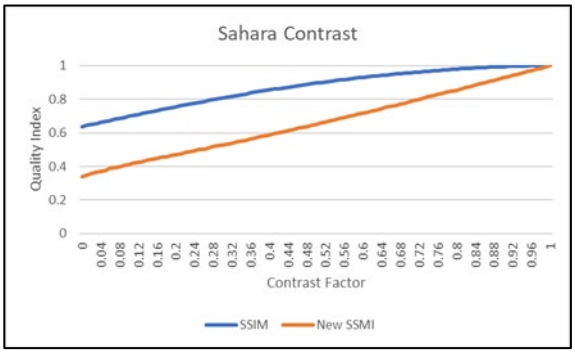
Unlike the pixel checkerboard, most of the actual image has different contrast levels because the complexity of the image is different, the amount of object or texture with a distinguishable intensity or colour is different between different images. Figure 9 shows images with different contrast factors of 0.2, 0.6 and 1.0. It can be seen that the image of Vietnamese street is a lot more complex compared to the image of the Sahara Desert, therefore, there is a lot of features for contrast in Vietnamese street compared to the Sahara Desert.

Figure 10 shows comparisons of images with different contrast with the image of contrast factor of 1. From the result, the characteristic of SSIM and new SSIM is similar to the earlier test where SSIM has a smooth slope and new SSIM is linear as the contrast factor increases to 1. The only difference is that how the results deviate as the image's contrast became closer to 0 compared with the original images. The quality index of SSIM for London, Vietnam and Sahara are 0.6901, 0.4440 and 0.7546 respectively, and the quality index of new SSIM is 0.3931, 0.2235 and 0.4703 at the contrast factor of 0.2. From these results, the average value of SSIM is 0.6296 and the average value of the new SSIM is 0.3623, at a contrast factor of 0.2. It is clear that SSIM is more permissive in its scoring system since the quality index is unreasonably high compared to the new SSIM.



(a)

(b)



(c)

Figure 10. Comparing image with different contrast with the image of contrast factor of 1. (a) London Bridge, (b) Vietnam Street, (c) Sahara Desert

The standard deviation of SSIM is 0.1338 and the standard deviation of the new SSIM is 0.1031 at a contrast factor of 0.2. The slightly larger standard deviation indicated that SSIM is a little bit inconsistent when comparing low contrast images compared to new SSIM. The main takeaway from this experiment is that the new SSIM able to tell the quality change on the image tested as the contrast factor change. The gradient of quality to contrast factor for new SSIM is almost the same across the whole contrast level on one image. Whereas, it is harder to correlate quality measurement of SSIM with the amount of change of contrast in an image.

3.3 Human Perception Test

The purpose of the human perception test is to examine the results of SSIM and new SSIM against the human evaluation of image quality. The test is done using the JPEG and JPEG2000 compression images, JPG compression is a popular method of reducing the storage size of images by reducing certain detail in the images. The procedure used in JPEG and JPEG2000 are different, therefore the compressed image also looks different. JPEG compressed seem to look more pixelated while JPEG2000 seem to look blurrier. This test used datasets provided by [13], the first dataset contained 29 high-resolution colour images that were compressed using JPEG at different compression ratios resulting in 204 compressed images and 29 original images, a total of 233 images. The second dataset also contained the same 29 high-resolution images but was compressed using JPEG2000 at different compression ratios resulting in 198 compressed images and 29 original images, a total of 227 images.

Test subjects were asked to provide their perception of the quality of images on a continuous linear scale and the scale was converted into a score of 1 to 100. The number of subjects for JPEG compressed is around 13 to 20 and for JPEG2000 is 25 subjects. The mean score for each image is taken as the representative score for that image from the subjective score files provided by the authors. In this experiment, the scores have not been processed and no outliers scores or subjects have been removed, which may explain why the data that appear in this test differ from the original papers. The average and median result of the subjective score for JPEG is 52.3433 and 57.75 respectively. The average and median result of JPEG2000 is 55.78037 and 61.08 respectively. The skewness of the total dataset is -0.392834686, the negative skewness indicates that most of the score concentrated on the higher side. But since the magnitude of skew is less than 0.5, most of the data is evenly or symmetrically distributed on both sides of the mean since the difference between mean and median are not too large for both JPEG and JPEG2000. The scatter plot of human subjective score versus model prediction of quality is shown in Figure 11.

The results of SSIM exponentially increase with the increment of human perception score. It has a large dispersion at the lower quality index, and it became smaller as the quality index become closer to 1. The scoring given by SSIM is also more generous compared to human perception in determining the quality of images since too much data is concentrated around the quality index of 1. As for the CW-SSIM, the curvature seems clearer compared to SSIM, hence the results seem to be exponentially related to the human perception. But the result appears to be more scattered and inconsistent among all 4. Similar to SSIM, it is easy to score high in the quality index using CW-SSIM because of how dense the data is around index 1 based on the graph.

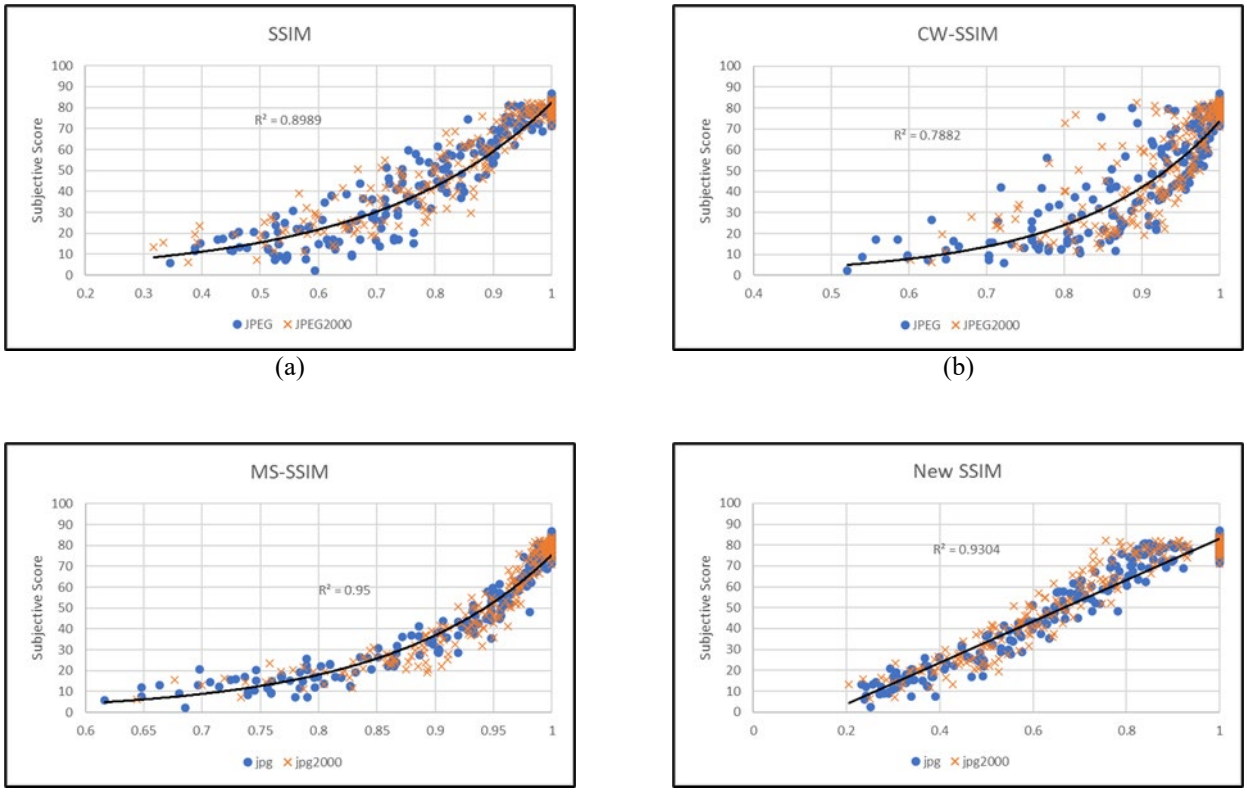


Figure 11. Scatter plot of mean subjective score versus model prediction on compressed images. (a) SSIM, (b) CW-SSIM, (c) MS-SSIM, (d) New SSIM

Table 1. R-squared and skewness of the scatter plot of each model. The best result is marked in red and the second-best result is marked in blue.

Model	R-squared	Skewness
SSIM	0.898941559	**-0.799095283**
CW-SSIM	0.788183158	-1.440498737
MS-SSIM	**0.949997465**	-1.393370212
New SSIM	**0.930352978**	**-0.258701243**

MS-SSIM is the most consistent scoring system based on how slim the data are compared to the others, it also indicates that it has a strong correlation with human perception. The exponential curve is also more obvious compared to SSIM and CW-SSIM. Even though it has a better result, the characteristic of SSIM where the data is highly crammed at high-quality index and loosely packed as the quality becomes lower is clearly visible. From the graph, the rate of change between new SSIM with human perception of quality is almost linearly related from the range of 0.1 to 0.8. The scatter becomes flat as the subjective score reach around 80 to 90. A small cluster that laying on the line of 1 and seem to be separated are the 29 uncompressed images of JPEG and 29 uncompressed of JPEG2000. This cluster can also be seen on all other graphs.

The characteristic of the new SSIM is different from the previous SSIM. All previous 3 SSIM tested in this paper share similar traits which are, the relationship of models with human perception is non-linear and the data is not properly distributed, the data gets denser as the quality index increases. R-squared is a measurement unit that show how close the scoring of each model fitted the regression line. The skewness of each result is also calculated in order to quantize the distribution of the scoring given by each model. The R-squared and skewness of SSIM, CW-SSIM, MS-SSIM and new SSIM are shown in Table 1.

The R-squared results show how strong the correlation between all SSIM scoring with the subjective score is. The regression line used for scattering data of SSIM, CW-SSIM and MS-SSIM is an exponential function since it has the highest R-squared result while new SSIM uses linear regression. With the highest R-squared, it is undoubtedly that MS-SSIM is the most consistent model of them all with new SSIM close second, followed by SSIM and CW-SSIM. The second variable that is also worth mentioning is skewness, it shows how the distribution for each model compared to the distribution of scores given by humans on the tested images. The skewness of human perception for both JPEG and JPEG2000 is -0.392834686. Skewness measures the asymmetry of the data distribution. The skewness of human scoring is negative with a value close to 0 meaning that the data is almost symmetrical with slight weighted to the high scoring. Since all SSIM results are negatively skewed and similar to human perception skewness meaning that all SSIM are in the correct distribution. CW-SSIM and MS-SSIM have an extreme negative skewness which highlights that both results are too concentrated on the high similarity even

though human scoring is almost symmetrical. The new SSIM has the most similar distribution to the human perception, followed by SSIM, MS-SSIM and CW-SSIM. Replacing the luminance and contrast component with the absolute difference equation drastically increase the linearity and consistency of new SSIM results with the human's average subjective score. The non-linear results of previous versions of SSIM make it harder for any study that creates image analysis models that are based on SSIM hoping for high resemblance with human perception

4. CONCLUSION

SSIM has been widely used in multiple fields of research especially in signal, image and video processing. However, the usage of mathematical equations in luminance and contrast may lead to counterintuitive results. Yes, the SSIM is capable of telling that the two images are similar, but it does not properly tell how different the two images are. From the experimental results, it is obviously shown that Equation (6) contributes to measurement bias, two high values are easier to score high similarity even though the difference between the two is clearly visible and when comparing two low values the score is low similarity even when the two images look almost the same. It was mentioned that that the luminance component of SSIM is qualitatively consistent with Weber's law of human perception of changes, but all the experiments on luminance stated otherwise. The judgement of the luminance component on intensity does not reflect human perception at all, it becomes overly sensitive on the low intensity and become unresponsive on the high intensity. Any model that used a luminance component for image assessment may be subjected to distortion on low intensity and can lead to undesired results. This may also apply to the contrast component as well since both components used the same equation.

There are many modifications and improvement that has been done for SSIM by various authors, but most of them still used the same mathematical relation. This biasing problem can be fixed by replacing the equation in the luminance and contrast with the relative difference equation. Since Weber's law stated that human perception of change in stimuli is proportional to background stimuli meaning that humans measure difference relative to background reference. We propose finding the absolute difference between intensity relative to the highest intensity as the new luminance component and using the ratio between low and high contrast on every pixel as the new contrast component. The modifications make the luminance and contrast response directly with the spatial distance between two signals, this makes the new SSIM overcome the distance function problem where the previous version lack. The proposed solution also does exceptionally well in comparing the quality between two images. Because of how similar it is compared to previous models with human perception also prove that human does measure the quality of image using relative difference more than using equation stated by original authors. The solution proposed is not the absolute answer for how humans measure the difference between images or the best image quality quantifier. It still has more room for improvements, and it is possible to add more component to it or merge it with other algorithms to make it more precise or sensitive.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Higher Education and a very special thanks go to the Research University Grant (RUG) of Universiti Teknologi Malaysia that has supported this research (project no. Q.J130000.4351.09G67).

REFERENCES

[1] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing*, 13(4), 2004, 600-612.

[2] Z. Wang, E. Simoncelli and A. Bovik, Multiscale structural similarity for image quality assessment, *The 37th Asilomar Conference on Signals, Systems & Computers*, California, USA, 2003.

[3] A. Bovik, Content-weighted video quality assessment using a three-component image model, *Journal of Electronic Imaging*, 19(1), 2010, 011003.

[4] Zhou Wang and A. Bovik, Mean squared error: Love it or leave it? A new look at signal fidelity measures, *IEEE Signal Processing Magazine*, 26(1), 2009, 98-117.

[5] F. Yan and C. Min, An improved method of SSIM based on visual regions of interest, *2015 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, Ningbo, China, 2015.

[6] R. Hassen and E. Steinbach, HSSIM: An objective haptic quality assessment measure for force-feedback signals, *Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, Cagliari, Italy, 2018.

[7] M. Aljanabi, Z. Hussain, N. Shnain and S. Lu, Design of a hybrid measure for image similarity: a statistical, algebraic, and information-theoretic approach, *European Journal of Remote Sensing*, 52(4), 2019, 2-15.

[8] E. Kandel, J. H. Schwartz and T. Jessell, *Principles of Neural Science*, Fifth edition, New York: McGraw-Hill Education, 2013.

[9] Statistics How To, *Correlation Coefficient: Simple Definition, Formula, Easy Steps*, 2021, https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/. (Accessed 31.12.2021).

[10] J. Terrace, M. McGreal and W. Morgenstern, *A Python module for computing the Structural Similarity Image Metric (SSIM)*, GitHub, 2021. https://github.com/jterrace/pyssim. (Accessed 31.12.2021).

[11] A. Khalel and S. Puranik, *All Image Quality Metrics You Need in One Package*, GitHub, 2021, https://github.com/andrewekhalel/sewar. (Accessed 31.12.2021).

[12] J. Nilsson and T. Akenine-Möller, *Understanding SSIM*, 2020. ArXiv, abs/2006.13846.

[13] H. R. Sheikh, Z. Wang, A. C. Bovik and L. K. Cormack, *Laboratory for Image and Video Engineering*, The University of Texas at Austin, 2021, http://live.ece.utexas.edu/research/quality/. (Accessed 31.12.2021).