A Formal Assessment of the Structural Similarity Index

Richard Dosselmann and Xue Dong Yang
Technical Report TR-CS 2008-2
September 2008

Department of Computer Science
University of Regina
Regina, Saskatchewan, Canada S4S 0A2

ISBN 0-7731-0649-9 (print)
ISBN 0-7731-0648-0 (on-line)
ISSN 0828-3494

A Formal Assessment of the Structural Similarity Index

Richard Dosselmann and Xue Dong Yang

Abstract

In recent years the structural similarity index has become a de facto standard among image quality metrics. Made up of three components, this technique assesses the visual impact of changes in luminance, contrast and structure in an image. Present applications of the index include image enhancement, video compression, video quality monitoring and signal encoding. As its status continues to rise, however, so do questions about its performance. In this paper, it is shown, both empirically and analytically, that the index is directly related to the conventional, and fallible, mean squared error. This relationship raises serious questions about the credibility of the structural metric.

Index Terms

structural similarity index, SSIM, mean squared error, MSE, image quality metric.

I. Introduction

Image quality assessment is an emerging field of signal processing. More or less defined as the task of designing an algorithm to automatically judge the perceived "quality" of a photograph, it remains a largely open problem. Though numerous algorithms, or *image quality metrics* [1], [2] (IQM), have been proposed, none truly correlates with the notion of "quality" as perceived by the *human visual system* [1] (HVS). Several attempts [3]–[5] have been made to simulate the intricate processes of the HVS. Most approaches, however, generally describe quality in terms of the pixel differences between an "original" image and its damaged, or *coded*, counterpart. For a given signal[1], its "original" form is one that is free of any distortions and is therefore assumed to be of perfect quality. Techniques that require both an original and coded image are known as *full reference* [2] metrics. *Reduced reference* [2] or *no reference* [2] metrics that require only a partial signal, or none at all, are usually more difficult to design. Both of the methods examined in this paper are full reference in nature. As well, each of the techniques is a pixel-difference "error" measure, as opposed to the more sophisticated "perceptual" type that incorporates advanced knowledge and properties of the HVS.

Not long ago, a new approach, based on statistical changes in image luminance, contrast and structure, was put forward. The method has quickly become the subject of considerable research and attention. The product of several years of research [6]–[10] itself, the metric has been formulated and revised a number of times. The procedure has even been extended to the complex wavelet domain [11]–[13]. It has been used in video quality monitoring [14]–[17], photographic restoration [18], biomedical imaging [19], imaging coding [20], video compression [21], [22], visual cognition [23] and picture enhancement [24]. Despite its numerous applications and growing body of followers, this research finds that the technique is both statistically and analytically linked to the most elementary and unreliable of all image quality metrics.

The two metrics considered in this investigation are introduced and discussed in Sections II and III, respectively. The structural metric is then thoroughly evaluated in Sections IV and V. Further analysis is presented in Section VI, before closing statements are made in Section VII.

II. Mean Squared Error

Perhaps no metric has received more attention than the *mean squared error* [1] (MSE). Its simple formulation and clear interpretation have allowed this technique to become the most widely employed metric in the field. Unfortunately, it often produces misleading values that do not correlate very well with perceived quality [1], [2], [6], [25], [26]. This fact is not especially surprising given that the metric is nothing more than a measure of the per-pixel differences between an original image and its distorted form.

A. Definition

Formally, the mean squared error is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2, \tag{1}$$

where n is the number of pixels in an n-dimensional image vector, and where x_i and y_i denote the i-th pixels of the original and coded image vectors $\mathbf{x}$ and $\mathbf{y}$, respectively. Mathematically, the MSE represents the average squared distance between two vectors, in this case, the image vectors $\mathbf{x}$ and $\mathbf{y}$.

Richard Dosselmann and Xue Dong Yang are with the Department of Computer Science, University of Regina, Regina, SK, S4S 0A2 Canada e-mail: {dosselmr, yang}@cs.uregina.ca.

[1]In this paper, the terms "image" and "signal" are used interchangeably.

(a) (b) (c) (d) (e) (f)

Fig. 1. Equal MSE Hypersphere. (a) Original "Barbara" image; (b) 3.25% random noise, MSE $= 253.5387$; (c) 11.25σ Gaussian noise, MSE $= 252.5258$; (d) 7×7 averaging blur, MSE $= 221.7118$; (e) 9×9 Gaussian blur, MSE $= 254.5792$; (f) 90% JPEG compression, MSE $= 250.0337$.

B. Equal MSE Hypersphere

One of the most frequently cited problems of the MSE is its inability to differentiate quality among different types of errors. For instance, although a picture damaged by random noise is normally of higher visual quality than one that is corrupted by a blur, the MSE, being highly susceptible to the large numerical deviations of the noise, incorrectly concludes that the former image is of lower quality than the latter. In order to reach the proper conclusion, the amount of blur must be dramatically increased. Regrettably, this problem extends far beyond noise and blur. In fact, for a variety of distortions, it is possible to generate a series of coded images that are nearly identical in terms of their measured MSE, yet are visually distinct in terms of their perceived quality. Arranging these images, equidistant, from a single original signal, one is able to construct an *equal MSE hypersphere* [10].

This concept is readily illustrated in the example of Fig. 1. in which each of the photos is approximately equidistant from the original, despite their striking differences in perceived quality. The first noisy image of Fig. 1b is quite close to the original, along with the second, given in Fig. 1c, which is only marginally affected by the superimposed noise. Conversely, the blurred and compressed pictures of Figs. 1d, 1e and 1f, respectively, are comparatively much worse. The compressed photo is particularly poor in this example. As this example plainly demonstrates, any gain that the mean squared error might offer in terms of simplicity is undoubtedly offset by its deplorable performance.

C. Error Map

An *error map* offers a means of visualizing the locations and magnitudes of the errors detected in a coded image. Here, darkly colored sections of the map, such as those in the background of Fig. 2c, represent areas in which there are few discernible errors. On the other hand, bright white portions indicate larger and often more noticeable impairments. In the specific case of the MSE, an error map is created by first calculating the individual error at a given pixel, namely the value $(x_i - y_i)^2$, and then subsequently normalizing it by the largest possible squared error, in this case the value 255^2. Note that in this particular implementation, all scores have additionally been stretched to the interval $[0, 255]$ to allow for proper display and viewing.

Example error maps showing the measured effects of noise, blur and JPEG compression are given in Figs. 2c, 2f and 2i, respectively. Because of the possible difficulties in viewing these particular error maps, they have each been enhanced by way of a *histogram equalization* [27], as seen in Figs. 2d, 2g and 2j. For the signal tainted by random noise, errors are sporadic, meaning that only those pixels directly impacted by noise appear in the resulting error map. A more coherent error map surfaces when a blur is present. This time, errors are most conspicuous along image contours, areas in which edges have been smeared. Smoother regions, however, are more or less preserved under the distortion. Compression produces a rather interesting error map. The averaging effect created by the compression scheme results in substantial errors not only along the 8×8 block boundaries, but also across individual blocks.

III. Structural Similarity Index

Given the obvious limitations of the mean squared error, Wang et al. [6]–[10], [14], [15] propose a more intelligent solution to the problem of image quality assessment. Made up of three terms, the *structural similarity* [9] (SSIM) index estimates the visual impact of shifts in image luminance, changes in photograph contrast, as well as any other remaining errors, collectively identified as structural changes. The metric is "based on a top-down assumption that the HVS is highly adapted for extracting structural information from the scene, and therefore a measure of structural similarity should be a good approximation of perceived image quality" [7]. According to its designers, the "SSIM correlates extraordinarily well with perceptual image quality, and handily outperforms prior state-of-the-art HVS-based metrics" [28]. In fact, "the degree of improvement obtained by SSIM relative to the prior standard-bearer is nearly equal to the progress made over the previous thirty years of research" [28]. As for the mean squared error, "the SSIM index has been shown to outperform MSE and the related PSNR in measuring the quality of natural images across a wide variety of distortions" [18]. Wang et al. contradict this last statement a number of times. Specifically, the graphical plots of Figs. 3, 5 and 7 of [7], [14] and [9], respectively, suggest that the performance of the PSNR[2] is just about identical to that of the SSIM. What's more, Rouse and Hemami find evidence of "a slight incompatibility

[2]The PSNR or *Peak signal-to-noise-ratio* is an adjusted form of the MSE. It is formally defined in [1].

Fig. 2. MSE error map. (a) Original "Lena" image; (b) 6.25% random noise; (c) MSE random noise error map; (d) Equalized MSE random noise error map; (e) 15×15 Gaussian blur; (f) MSE Gaussian blur error map; (g) Equalized MSE Gaussian blur error map; (h) 85% JPEG compression; (i) MSE JPEG compression error map; (j) Equalized MSE JPEG compression error map.

between signal structure, as given by the structure component of MSSIM, and perceived visual structure" [23]. A formal definition of the MSSIM is given in the next section. And when it comes to video quality, Vorren "cannot statistically prove that SSIM is better, due to the overlapping of the confidence intervals of all the video quality models" [29]. The results presented in this paper also contradict the findings of Wang et al. In fact, far from superior, the SSIM index is found to be directly linked with the mean squared error.

A. Definition

For original and coded signals $\mathbf{x}$ and $\mathbf{y}$, respectively, the SSIM index is defined as

$$SSIM(\mathbf{x},\mathbf{y}) = [l(\mathbf{x},\mathbf{y})]^{\alpha}[c(\mathbf{x},\mathbf{y})]^{\beta}[s(\mathbf{x},\mathbf{y})]^{\gamma}, \tag{2}$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ control the relative significance of each of the three terms of the index. In this implementation, like that of Wang et al. [9], $\alpha = \beta = \gamma = 1$. The formulation given in (2) satisfies a number of properties. First, it guarantees *symmetry* [9], meaning that $SSIM(\mathbf{x},\mathbf{y}) = SSIM(\mathbf{y},\mathbf{x})$. Thus, the original and coded images may be swapped. As well, the metric ensures *boundness* [9], in the sense that $SSIM(\mathbf{x},\mathbf{y}) \leq 1$. Finally, there is a *unique maximum* [9], meaning that $SSIM(\mathbf{x},\mathbf{y}) = 1$ if and only if $\mathbf{x} = \mathbf{y}$. The luminance, contrast and structural components of the index are defined individually as

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{3}$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{4}$$

(a) (b) (c) (d) (e) (f)

Fig. 3. Equal SSIM Hypersphere. (a) Original "Barbara" image; (b) 7.15% random noise, SSIM = 0.5033, MSE = 550.1926; (c) 14.5σ Gaussian noise, SSIM = 0.5036, MSE = 418.5383; (d) 9×9 averaging blur, SSIM = 0.4989, MSE = 279.0549; (e) 13×13 Gaussian blur, SSIM = 0.4767, MSE = 369.0773; (f) 90% JPEG compression, SSIM = 0.5028, MSE = 250.0337.

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}, \tag{5}$$

where μ_x and μ_y represent the means of the original and coded images, respectively, σ_x and σ_y represent the standard deviations of each of the signals and σ_{xy} is the covariance of the two photos. As a means of dealing with the situations in which the denominators are close to zero, the constants C_1, C_2 and C_3 are introduced. For an 8-bit grayscale image composed of $L = 2^8 = 256$ gray-levels, $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$ and $C_3 = C_2/2$, where $K_1 = 0.01$ and $K_2 = 0.03$. When $C_1 = C_2 = 0$, the metric is reduced to the *universal quality index* [8] (UQI), a predecessor of the SSIM index. Note that (5), absent the C_3 constant, represents the *linear correlation* [30] of the two signals. Unlike the MSE, which is measured at the global level, the SSIM is computed locally. An 8×8 window moves, a single pixel at a time, across an image. At each step, a local SSIM score is calculated. The final score of an entire photograph, referred to as the *mean SSIM* [9] (MSSIM)[3], is the simple arithmetic average of each of the local scores.

At this time, the SSIM does not directly incorporate any advanced features of the human visual system, despite assertions by the group that the metric is based on a "*top-down* approach, mimicking the hypothesized functionality of the overall HVS" [9]. Almost contradicting this statement, the team finds it "surprising that such a simple mathematically defined quality index performs so well without any HVS model explicitly employed" [8]. Perhaps a bit too simple, Beghdadi and Iordache note that "the SSIM does not take into account the frequency content of the image which plays an important role in the discrimination between spatial structures" [31].

B. Equal SSIM Hypersphere

Wang et al. routinely boast about the ability of their metric to correctly distinguish perceived quality over different kinds of distortions in situations in which the mean squared error is unable to do the same. The group concludes that "images with similar MSE have significantly different visual quality, and MSSIM delivers better consistency with perceptual evaluations" [14]. Ironically, there are times in which the opposite is true. Consider the following example of Fig. 3. Like the MSE hypersphere of Section II-B, each of the coded images is at a nearly equal distance from the original, this time in terms of its measured SSIM score. Just as the MSE does, the SSIM erroneously concludes that the noisy images of Figs. 3b and 3c are visually equivalent to the blurred and compressed signals of Figs. 3d, 3e and 3f. The observant reader will spot the similarities between the images in this example and those of the MSE of Fig. 1. With similar conclusions over a common set of test images, the parallels between the structural index and MSE are growing ever more obvious. This research is not the first to unearth such problems. The index is also seen to incorrectly order images in Figs. 3 and 4 of [31].

As an added note on this subject, it is interesting to observe that Wang et al. appear to only ever consider one particular scenario when constructing an equal MSE hypersphere. In Figs. 1, 2, 2, 3, 4 and 41.9 of [11], [8], [9], [6], [10] and [15], respectively, a given photograph is repeatedly subjected to the same set of impairments. First, signal luminance and contrast are altered. Next, a slight touch of random noise is added, presumably since the SSIM, like the MSE, appears to be rather sensitive to the effects of noise. A somewhat more destructive blur is then introduced to produce a fourth coded image. In the last step, the picture is heavily compressed with JPEG, resulting in a fifth image that is highly corrupted, both visually and numerically. This particular set-up is precisely that given in this paper in Fig. 4, not by chance, but rather by necessity. In fact, it is difficult to construct any other situation involving these five distortions, in which the individual MSE scores are approximately equal, while the SSIM values are more or less correct. It seems that this contrived scenario is one of only a few that actually results in the desired outcome. The secret behind this ruse lies in the fact that the MSE is highly sensitive to changes in luminance and contrast. Though visually insignificant to a human observer, such changes result in considerable numerical deviations when measured by the mean squared error. From here, one is then able to create noisy, blurred and compressed images with equally

[3]In this paper, the terms SSIM and MSSIM are used interchangeably. Both are assumed to represent a single quality score that is the arithmetic average of a series of local scores carried out over individual 8×8-sized blocks.

(a) (b) (c) (d) (e) (f)

Fig. 4. Equal MSE Hypersphere. (a) Original "Barbara" image; (b) 7.45% luminance shift, SSIM $= 0.9815$, MSE $= 361.0000$; (c) 30% contrast adjustment, SSIM $= 0.8689$, MSE $= 398.7237$; (d) 5.15% random noise, SSIM $= 0.5839$, MSE $= 370.4489$; (e) 13×13 Gaussian blur, SSIM $= 0.4767$, MSE $= 369.0773$; (f) 99% JPEG compression, SSIM $= 0.3861$, MSE $= 384.2890$.

large error scores. More importantly, these three images appear visibly corrupted to the viewer, such as those of Figs. 4e and 4f. The final product is one in which the first two images, namely those of Figs. 4b and 4c, are of near-perfect quality, while the remaining three, given in Figs. 4d, 4e and 4f, are visibly marred. Because the structural index handles variations in luminance and contrast separately, the overall SSIM score is not particularly affected by such changes. This realization allows one to hold down the value of the SSIM in the first two images, while freely adjusting its score over the remaining photos. Thus, in some sense, the luminance and contrast terms serve as a buffer. Once removed, however, the SSIM rises and falls in much the same way as the MSE, as demonstrated in Fig. 3, exposing the deception. Moreover, attempts to further modify the last three images of Fig. 4, by injecting additional noise for instance or decreasing the amount of the blur, result in incorrect scores on the part of the SSIM. Additional noise, to which the SSIM is very sensitive, quickly brings down the overall quality score of the noisy image. Consequently, the noisy image is improperly assigned a lower quality score than the blurred and compressed images. In the opposite way, decreasing the amount of blur or compression causes the scores of the last two images to incorrectly rise above that of the noisy image. Before closing this section, note that for a contrast adjustment or equalization, the SSIM, like the MSE, improperly assigns the enhanced photograph a lower quality score than the original. As an example of this contradiction, see Fig. 4c above.

C. Error Map

Similar to the MSE error map of Section II-C, an SSIM error map may also be generated. In much the same way, it visually highlights the areas of an image in which errors are located. Each pixel in this particular map denotes the local SSIM score computed using the neighboring pixels of the 8×8 local image region. Because an SSIM score does not exceed 1, there is no need for any normalization. Instead, the individual values are simply scaled to the interval $[0, 255]$ to allow for suitable viewing.

Figs. 5c, 5f, 5i illustrate the behavior of the SSIM metric when noise, blur and compression are present. Equalized forms of the maps, though not particularly needed in this case, are given in Figs. 5d, 5g, 5j. The noise map of Fig. 5c is somewhat different from that of the MSE, given in Fig. 2c. Whereas the MSE map emphasizes only those pixels in which noise actually occurs, the SSIM, given its local formulation, takes into account nearby adjoining pixels. In some sense, the SSIM error map is a blurred form of that of the MSE. The differences are much smaller in the case of a blur. One notices that the SSIM map of Fig. 5f appears much brighter than that of the MSE of Fig. 2f. However, when compared with the equalized MSE map of Fig. 2g, the distinctions are much less pronounced. Both show significant errors in the region descending from hat, with lesser amounts in the smooth portions of the shoulder, mirror and back wall. Again, the SSIM map is much like a blurred version of the MSE error map. Finally, there are some discrepancies between the compression maps of Figs. 5i and 2i. It is possible, however, to produce an error map much like that of Fig. 5i, by simply computing the MSE locally over individual 8×8 blocks. This finding suggests that the differences in the various error maps are likely the result of nothing more than the decision to apply the SSIM metric locally rather than globally. Interestingly, the SSIM map of Fig. 5i indicates that there are no major errors along image contours, even though many, such as those along the edges of the mirror and hat, have been noticeably warped by the compression scheme.

IV. Empirical Assessment

In this first of two evaluations of the SSIM index, the scoring behavior of the metric is statistically compared[4] with that of the MSE. Ensuing *regression analysis* [32] suggests an almost perfect correlation between the two.

A. Method

Prior to the experiment, a set of 15 test images was put together. Each of the 15 test images was then synthetically corrupted using 7 types of errors, resulting in $15 \times 7 = 105$ data sets. The specific impairments considered in this study are those of noise,

[4]All data was analyzed using SAS® 9.1 statistical software.

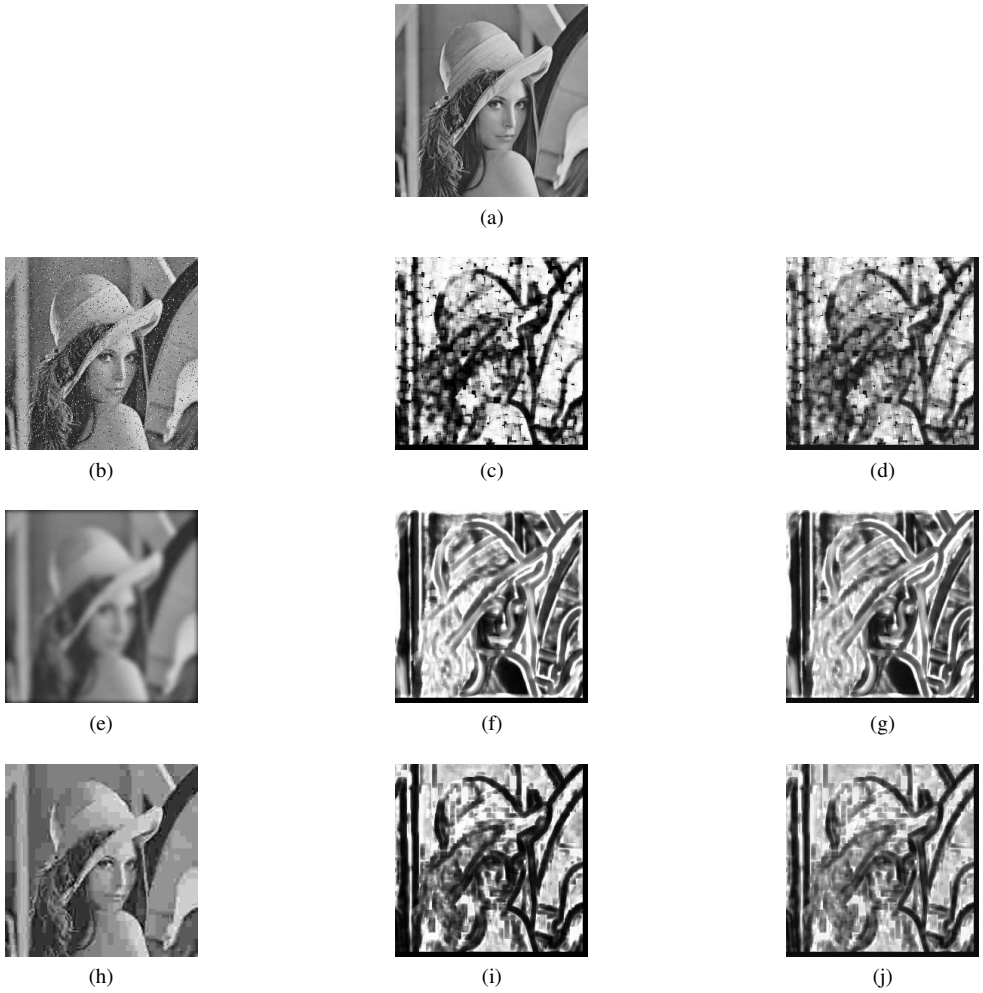

Fig. 5. SSIM error map. (a) Original "Lena" image; (b) 6.25% random noise; (c) SSIM random noise error map; (d) Equalized SSIM random noise error map; (e) 15×15 Gaussian blur; (f) SSIM Gaussian blur error map; (g) Equalized SSIM Gaussian blur error map; (h) 85% JPEG compression; (i) SSIM JPEG compression error map; (j) Equalized SSIM JPEG compression error map.

blur, JPEG compression, luminance shifts and contrast adjustments. Next, for each of the 7 types of errors, 20 individual coded images were generated by varying the amount of distortion introduced. Collectively, this resulted in 7 sets of 20 corrupted images for each of the 15 test images. Then, within the 105 data sets, quality scores for each of the 20 images were obtained by comparing the given pictures with their original forms using both the SSIM and MSE metrics. Later, using *quadratic regression* [32], the resulting 20 SSIM scores were regressed on the 20 MSE scores. These relationships are represented by 105 individual quadratic models, each of the form

$$MSE = \beta_0 + \beta_1 SSIM + \beta_2 SSIM^2, \tag{6}$$

where β_0, β_1 and β_2 denote the intercept and slope parameters of each of the 105 models. These parameters vary between data sets because of the different means and variances of the individual test images. In the final step, the degree of correlation between the SSIM and MSE in each of the 105 models was measured using R^2, the *coefficient of multiple determination* [32]. Much like linear correlation, an R^2 value is given on the interval $[0, 1]$, where a score near 0 indicates little association between the two variables, while a value close to 1 is evidence of a stronger degree of correspondance.

B. Results and Discussion

Individual R^2 correlations for each of the 105 models are found in Table I. The exceptionally high level of correlation between the SSIM and MSE is obvious. Values range from $R^2 = 0.9322$ to $R^2 = 1.0000$, with an average of $R^2 = 0.9867$. Shifts in luminance result in the strongest associations, near perfect at times. Relationships involving random noise, Gaussian noise and JPEG compression are also decidedly sound. Results this compelling are no doubt a sign of a deeper relationship between the two methods.

Seven of the quadratic-shaped relationships are graphically depicted in Fig. 6. Curves are remarkably smooth, with no major outliers. These virtually perfect relationships are a further indication of a functional dependency. Minor perturbations

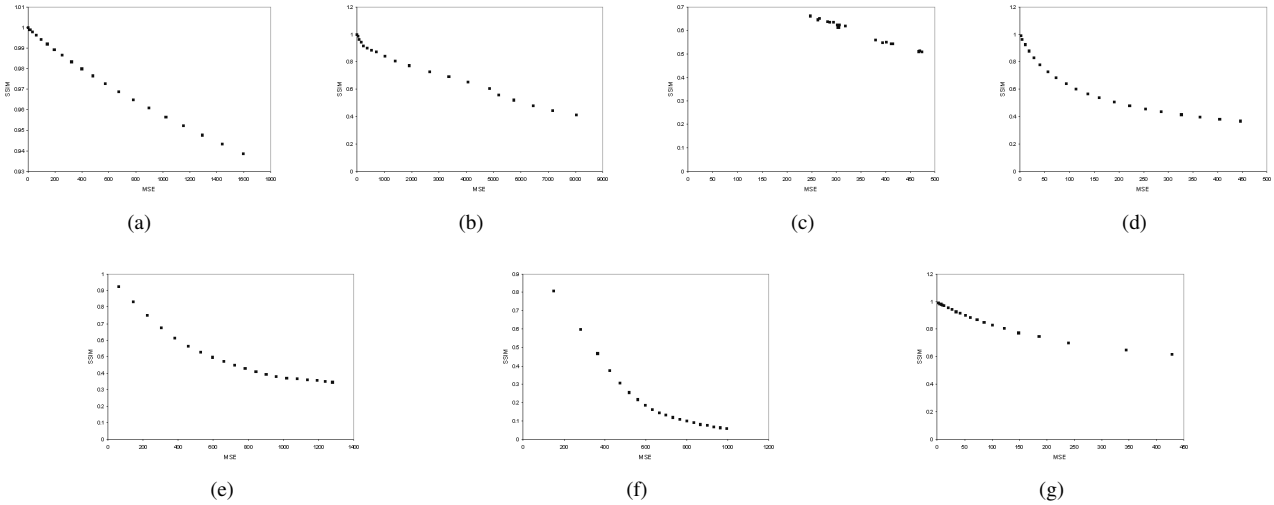

Fig. 6. MSE vs. SSIM for seven selected models. (a) Luminance shift ("Barbara" image); (b) Contrast adjustment ("boat" image); (c) Random noise ("Einstein" image); (d) Gaussian noise ("house" image); (e) Averaging blur ("Lena" image); (f) Gaussian blur ("mandrill" image); (g) JPEG compression ("peppers" image).

TABLE I
R^2 CORRELATION COEFFICIENTS OF MSE VS. SSIM

Test Image	**Luminance Shift**	**Contrast Adjustment**	**Random Noise**	**Gaussian Noise**	**Averaging Blur**	**Gaussian Blur**	**JPEG Compression**	**AVERAGE**
Barbara	1.0000	0.9847	0.9957	0.9975	0.9780	0.9906	0.9957	0.9917
boat	1.0000	0.9939	0.9910	0.9926	0.9820	0.9808	0.9956	0.9908
cameraman	0.9992	0.9710	0.9922	0.9832	0.9812	0.9757	0.9792	0.9831
couple	0.9999	0.9875	0.9960	0.9978	0.9362	0.9744	0.9973	0.9842
Einstein	1.0000	0.9487	0.9907	0.9952	0.9754	0.9798	0.9959	0.9837
Goldhill	1.0000	0.9727	0.9971	0.9977	0.9657	0.9645	0.9965	0.9849
house	1.0000	0.9562	0.9930	0.9816	0.9633	0.9846	0.9862	0.9807
lake	0.9985	0.9808	0.9945	0.9951	0.9549	0.9821	0.9991	0.9864
Lena	1.0000	0.9838	0.9935	0.9924	0.9722	0.9932	0.9959	0.9901
man	1.0000	0.9903	0.9948	0.9973	0.9414	0.9813	0.9966	0.9860
mandrill	1.0000	0.9897	0.9963	0.9994	0.9322	0.9506	0.9950	0.9805
MIT	0.9999	0.9499	0.9879	0.9926	0.9698	0.9921	0.9943	0.9838
peppers	1.0000	0.9856	0.9892	0.9925	0.9738	0.9912	0.9955	0.9897
Tiffany	1.0000	0.9987	0.9980	0.9947	0.9798	0.9921	0.9978	0.9944
woman	1.0000	0.9838	0.9903	0.9872	0.9897	0.9836	0.9994	0.9906
AVERAGE	0.9998	0.9785	0.9933	0.9931	0.9664	0.9811	0.9947	0.9867

are detected in Figs. 6b and 6c. These slight deviations are attributed to the fact that changes in luminance and contrast are not entirely independent of those in structure. Likewise, adjustments to image structure indirectly affect luminance and contrast. Wang et al. indicate that "the three components are relatively independent" [10], meaning that this interaction is minimal, though numerically it is apparent in Table I. To better understand this occurrence, consider, for example, the situation in which the dynamic range of an already high-contrast image is expanded. Pixels near the discrete upper and lower bounds, 0 and 255, respectively, are only marginally affected by the transformation. Pixels in the middle range, conversely, are more likely to be further spread out. This non-linear transformation alters the relative ordering of pixel values, thereby resulting in a structural change in addition to the adjustment in signal contrast. It is this phenomenon that is responsible for the minor variations that are observed in Fig. 6b, as well as the third column of Table I. As for Fig. 6c, these fluctuations are due to tiny modifications in luminance and contrast resulting from the addition of noise, changes which affect the structural component. Similar alterations to luminance and contrast are produced by blurring and compression. On the whole, contrast appears to be far more sensitive to structural changes than luminance. Even the addition of small quantities of noise, blur or compression causes the level of correlation to drop below 1. Luminance, conversely, is more independent. As proof, one need only consider almost perfect correlation of the second column of Table I. Had the three terms of the index been completely independent of another, there would likely have been perfect correlation across the board.

V. Formal Assessment

One is motivated by the empirical findings of the preceding section to formally prove that the structural index and mean squared error are indeed connected. In the following sections, the two are algebraically shown to be functions of one another, thereby eliminating any doubt about their association. Verification of correspondence is straightforward. Definitions of the mean squared error, along with each of the three components of SSIM are expanded, and then, term by term, are linked.

A. Luminance Term

A basic increase or decrease in image luminance is the first issue to be addressed. Here, a signal is altered only by the addition of a scalar value, identified as δ. The relative ordering of pixel values is preserved, meaning that there is no adjustment in contrast. In the same way, if the addition of δ does not cause a pixel to go beyond the discrete boundaries, again the values 0 and 255, respectively, then there will be no change in structure either.
Expanding the mean squared error, specifically the definition of (1), one finds that

$$\begin{aligned} MSE &= \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i y_i + \frac{1}{n}\sum_{i=1}^{n}y_i^2 && (7) \\ &= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i(x_i+\delta) + \frac{1}{n}\sum_{i=1}^{n}(x_i+\delta)^2 && (8) \\ &= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i^2 - \frac{2\delta}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{2\delta}{n}\sum_{i=1}^{n}x_i + \delta^2 && (9) \\ &= \delta^2. && (10) \end{aligned}$$

The MSE is therefore reduced to the square of the change in luminance, namely δ^2. Next, the luminance term of (3) is expanded as

$$\begin{aligned} l(\mathbf{x},\mathbf{y}) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \\ &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + (\mu_x+\delta)^2 + C_1} && (11) \\ &= \frac{2\mu_x\mu_y + C_1}{2\mu_x^2 + 2\delta\mu_x + \delta^2 + C_1} && (12) \\ &= \frac{2\mu_x\mu_y + C_1}{2\mu_x(\mu_x+\delta) + \delta^2 + C_1} && (13) \\ &= \frac{2\mu_x\mu_y + C_1}{2\mu_x\mu_y + \delta^2 + C_1}. && (14) \end{aligned}$$

Replacing the δ^2 term by the mean squared error, one is able to connect the MSE and luminance terms. In particular,

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{2\mu_x\mu_y + MSE + C_1}. \quad (15)$$

The reverse form, in which the MSE is derived from the luminance component, is given as

$$MSE = \frac{2\mu_x\mu_y + C_1}{l(\mathbf{x},\mathbf{y})} - 2\mu_x\mu_y - C_1. \quad (16)$$

In the first equation, namely (15), the MSE occurs in the denominator, resulting in an inverse, quadratic-like relationship, much like that of Fig. 6a. The same is true of (16), in which the luminance term also appears in the denominator.

B. Contrast Term

At present, there is no universal definition of a contrast change. Whereas an elementary technique might simply spread out pixel values, the approach taken in this paper, a more elaborate mechanism might equalize or redistribute pixels over the range of gray-levels. With no explicit definition of such a change, it is not possible to derive a general relationship to link the SSIM and MSE in this respect. Consequently, the mapping functions are given simply as

$$MSE = f_{c_1}(c(\mathbf{x}, \mathbf{y}), C_2), \tag{17}$$

and

$$c(\mathbf{x}, \mathbf{y}) = f_{c_2}(MSE, C_2). \tag{18}$$

Experimental evidence in Section V-D suggests that in most instances this term may be safely ignored.

C. Structural Term

A breakdown in image structure is the most complicated of the three scenarios. In this research, a structural modification is interpreted as any change in the makeup of signal that is not a simple shift in luminance or adjustment of contrast.

The MSE is again expanded as

$$\begin{aligned} MSE &= \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i y_i + \frac{1}{n}\sum_{i=1}^{n}y_i^2 \qquad (19) \\ &= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{1}{n}\sum_{i=1}^{n}y_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i y_i. \qquad (20) \end{aligned}$$

The structural term of (5) is given as

$$\begin{aligned} s(\mathbf{x}, \mathbf{y}) &= \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \\ &= \frac{\frac{1}{n}\sum_{i=1}^{n}x_i y_i - n\mu_x\mu_y + C_3}{\sqrt{\left(\sum_{i=1}^{n}x_i^2 - n\mu_x^2\right)\left(\sum_{i=1}^{n}y_i^2 - n\mu_y^2\right)} + C_3}. \end{aligned} \tag{21}$$

Adding several terms to (20) gives

$$\begin{aligned} \frac{MSE + 2(m_1 - C_3)}{d + C_3} &= \frac{\frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{1}{n}\sum_{i=1}^{n}y_i^2}{d + C_3} - \frac{2\left(\frac{1}{n}\sum_{i=1}^{n}x_i y_i - m_1 + C_3\right)}{d + C_3} \qquad (22) \\ &= \frac{\frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{1}{n}\sum_{i=1}^{n}y_i^2}{d + C_3} - 2s(\mathbf{x}, \mathbf{y}), \qquad (23) \end{aligned}$$

where $d = \sqrt{\left(\sum_{i=1}^{n}x_i^2 - n\mu_x^2\right)\left(\sum_{i=1}^{n}y_i^2 - n\mu_y^2\right)}$ and $m_1 = n\mu_x\mu_y$.

Thus, knowing the MSE, the structural term is computed as

$$s(\mathbf{x}, \mathbf{y}) = \frac{\frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{1}{n}\sum_{i=1}^{n}y_i^2 - MSE - 2(m_1 - C_3)}{2(d + C_3)}. \tag{24}$$

In a likewise manner, the MSE is found to be

1	2	3	4	5	6	7	8	8	8	8	8	8	8	8	8	8	7	6	5	4	3	2	1
2	4	6	8	10	12	14	16	16	16	16	16	16	16	16	16	16	14	12	10	8	6	4	2
3	6	9	12	15	18	21	24	24	24	24	24	24	24	24	24	24	21	18	15	12	9	6	3
4	8	12	16	20	24	28	32	32	32	32	32	32	32	32	32	32	28	24	20	16	12	8	4
5	10	15	20	25	30	35	40	40	40	40	40	40	40	40	40	40	35	30	25	20	15	10	5
6	12	18	24	30	36	42	48	48	48	48	48	48	48	48	48	48	42	36	30	24	18	12	6
7	14	21	28	35	42	49	56	56	56	56	56	56	56	56	56	56	49	42	35	28	21	14	7
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
8	16	24	32	40	48	56	64	64	64	64	64	64	64	64	64	64	56	48	40	32	24	16	8
7	14	21	28	35	42	49	56	56	56	56	56	56	56	56	56	56	49	42	35	28	21	14	7
6	12	18	24	30	36	42	48	48	48	48	48	48	48	48	48	48	42	36	30	24	18	12	6
5	10	15	20	25	30	35	40	40	40	40	40	40	40	40	40	40	35	30	25	20	15	10	5
4	8	12	16	20	24	28	32	32	32	32	32	32	32	32	32	32	28	24	20	16	12	8	4
3	6	9	12	15	18	21	24	24	24	24	24	24	24	24	24	24	21	18	15	12	9	6	3
2	4	6	8	10	12	14	16	16	16	16	16	16	16	16	16	16	14	12	10	8	6	4	2
1	2	3	4	5	6	7	8	8	8	8	8	8	8	8	8	8	7	6	5	4	3	2	1

Fig. 7. 24×24 coefficient map.

$$MSE = \frac{1}{n}\sum_{i=1}^{n} x_i^2 + \frac{1}{n}\sum_{i=1}^{n} y_i^2 - 2(s(\mathbf{x},\mathbf{y})(d + C_3) + m_1 - C_3). \tag{25}$$

Before continuing, define $l_1 = \sum_{i=1}^{n} y_i^2$.

Initially, (24) and (25) might look rather complicated. Closer inspection reveals that the two ultimately differ only by the term l_1. Assuming that the coded image remains unchanged under the distortion, as is normally the situation, the remaining terms of (24) and (25) are constants. As the amount of impairment increases, the magnitude of l_1 rises in a quadratic fashion. This factor accounts for the quadratic shape of the curves in Fig. 6. Evidently, the only "gain" offered by the SSIM over the MSE is the addition of the l_1 term, a simple proportionate measure of the amount of error occurring in a coded signal. Without this particular term, the two are equal, up to a constant value. Moreover, as the level of error increase, l_1 dominates, meaning that the SSIM becomes little more than a raw measure of error.

D. Conversion Algorithms

Since an SSIM score is computed locally, whereas an MSE is carried out at the global scope, a few processing steps, in addition to the formulas of the proceeding section, are required during conversion. Luckily, a straightforward modification switches a global MSE to a local measure. Computing the MSE over local 8×8 neighbors, each pixel in the interior region of the image is processed 64 times. Each pixel is also divided by 64 as part of each local 8×8 average. In total, these interior points are each processed one time. Thus, when collectively averaged, the average of the local mean squared errors is almost equal to the single global definition of (1). As illustrated in the 24×24-sized example image of Fig. 7, each term in the interior portion of an image is considered 64 times as the 8×8 window passes over. Coefficients along the boundaries, conversely, reflect the fact that these pixels are each processed less than 64 times. The solution is therefore to merely add to the average of the local MSE scores, the missing terms along the borders. Specifically, for each pair of boundary pixels x_i and y_i, actually processed a total of c_n times, where $1 \leq c_n < 64$, the value

$$\left(\frac{64c_n}{64}\right)(x_i - y_i)^2 \tag{26}$$

is added to the average. This adjusted average is than divided by n, resulting in the definition of (1). In general, the contribution of the boundary pixels rapidly begins to fall as the size of the image increases.

The following algorithm allows one to compute the mean squared error of a given image using the SSIM and the formulas of this section. Note that MSE^* refers to the local MSE computed in each 8×8 block.

$MSE \leftarrow 0$
$num\_local\_windows \leftarrow 0$

For each local 8×8 window
    $SSIM \leftarrow$ (2)
    $MSE^* \leftarrow$ (16), (17) or (25), via $SSIM$
    $MSE \leftarrow MSE + MSE^*$
    $num\_local\_windows \leftarrow num\_local\_windows + 1$

TABLE II
COMPUTING MSE FROM SSIM AND VICE VERSA, UNDER A SHIFT IN LUMINANCE.

ΔL	MSE	SSIM	SSIM to MSE	MSE to SSIM
5	25.0000	0.9984	25.0000	0.9984
10	100.0000	0.9941	100.0000	0.9941
15	225.0000	0.9878	225.0000	0.9878
20	400.0000	0.9798	400.0000	0.9798
25	624.9971	0.9707	624.9970	0.9707
30	899.9851	0.9606	899.9836	0.9606
35	1224.9103	0.9498	1224.9031	0.9498
40	1599.6284	0.9385	1599.6010	0.9385
45	2023.6811	0.9267	2023.5965	0.9268
50	2495.9745	0.9139	2495.7406	0.9144

End For

$MSE \leftarrow MSE/num\_local\_windows$
$MSE \leftarrow MSE + f_b(\mathbf{x}, \mathbf{y})$

The function $f_b(\mathbf{x}, \mathbf{y})$ adjusts the final total to include the missing pixels along the image perimeter.

The SSIM is calculated from the MSE via the ensuing steps.

$SSIM \leftarrow 0$
$num\_local\_windows \leftarrow 0$

For each local 8×8 window
  $MSE^* \leftarrow$ (1)
  $SSIML \leftarrow$ (15), via MSE^*
  $SSIMC \leftarrow$ (18), via MSE^*
  $SSIMS \leftarrow$ (24), via MSE^*
  $SSIM \leftarrow SSIM +$ (2),
    via $SSIML$, $SSIMC$ and $SSIMS$
  $num\_local\_windows \leftarrow num\_local\_windows + 1$
End For

$SSIM \leftarrow SSIM/num\_local\_windows$

Four test cases, given in Tables II, III, IV and V, demonstrate the conversion between the two metrics. Results are generally better when converting from SSIM to MSE, as opposed to the other direction. This fact is attributed to the absence of the contrast term. Still, these results prove that one is able to obtain the MSE from the SSIM. Table II, shows the effects of a variation in luminance. The amount of the specific change is given by ΔL. Conversion in either direction is just about perfect, given that that modification of the luminance term does not generally impact the other two. Unfortunately, the introduction of Gaussian noise, a Gaussian blur and JPEG compression alters each of the elements of the structural index. Alterations of contrast are not considered, given that no single conversion method exists in this case. Note that σ_1 controls the amount of Gaussian noise, σ_2 is the radius of the Gaussian blurring filter and J denotes the level of JPEG compression. Each table involves a separate test, in particular "Barbara", "Einstein", "Lena" and "Tiffany", respectively.

VI. FURTHER ASSESSMENT

A couple of remaining items of discussion are included in the following segments.

A. Error-Based SSIM

This investigation further advances the idea of the SSIM index as a trivial error-based quality metric, rather than a complex perceptual method. If the empirical findings of Section IV are not already proof enough, then surely the mathematical relationships of Section V confirm this hypothesis. In either case, the SSIM is consistently associated with the MSE, a technique devoid of any sophisticated reasoning. Peering deeper, it is clear that the structural index is nothing more than a comparison of signal means, variances and covariance. Thus, there is little cause to believe that it would ultimately offer any sizeable gains over the inept mean squared error. Nevertheless, Wang et al. declare the luminance term of (3) to be "qualitatively consistent with Weber's law, which has been widely used to model light adaptation (also called luminance masking) in the

TABLE III
COMPUTING MSE FROM SSIM AND VICE VERSA, UNDER THE ADDITION OF GAUSSIAN NOISE.

σ_1	**MSE**	**SSIM**	**SSIM to MSE**	**MSE to SSIM**
1.50	4.9808	0.9492	4.9808	0.9527
3.00	19.2015	0.8635	19.2015	0.8774
4.50	42.0390	0.7771	42.0390	0.8027
6.00	73.1636	0.6993	73.1636	0.7353
7.50	113.8126	0.6263	113.8126	0.6732
9.00	163.0459	0.5625	163.0459	0.6178
10.50	219.9483	0.5099	219.9483	0.5731
12.00	286.5716	0.4628	286.5716	0.5323
13.50	361.7418	0.4201	361.7418	0.4933
15.00	441.8168	0.3846	441.8168	0.4615

TABLE IV
COMPUTING MSE FROM SSIM AND VICE VERSA, UNDER THE EFFECTS OF A GAUSSIAN BLUR.

σ_2	**MSE**	**SSIM**	**SSIM to MSE**	**MSE to SSIM**
3×3	45.0053	0.9122	45.0053	0.9325
5×5	114.0283	0.8035	114.0283	0.8477
7×7	182.6360	0.7061	182.6360	0.7701
9×9	250.6562	0.6188	250.6562	0.6969
11×11	317.7825	0.5423	317.7825	0.6307
13×13	383.5542	0.4757	383.5542	0.5713
15×15	448.0828	0.4190	448.0828	0.5191
17×17	510.5293	0.3712	510.5293	0.4733
19×19	570.6115	0.3307	570.6115	0.4333
21×21	628.3643	0.2960	628.3643	0.4008

TABLE V
COMPUTING MSE FROM SSIM AND VICE VERSA, WITH THE INTRODUCTION OF JPEG COMPRESSION.

J	**MSE**	**SSIM**	**SSIM to MSE**	**MSE to SSIM**
10	8.1999	0.9377	8.1999	0.9387
20	16.8814	0.9055	16.8814	0.9071
30	31.7243	0.8540	31.7243	0.8584
40	50.3507	0.7927	50.3507	0.8070
50	68.4726	0.7432	68.4726	0.7654
60	88.5658	0.6885	88.5658	0.7422
70	115.6789	0.6353	115.6789	0.7159
80	154.3139	0.5741	154.3139	0.6906
90	223.5573	0.4809	223.5573	0.6778
99	354.5098	0.3663	354.5098	0.6328

HVS" [9]. *Weber's law* [1] expresses the relationship between the physical amount of light present versus the level actually perceived by a human observer. If the SSIM is indeed able to model this behavior, then apparently so is the MSE. One will recall the almost flawless correlation among the two methods during a shift in luminance. The findings of this research also cast doubts on claims that the contrast term of (4) "is consistent with the contrast-masking feature of the HVS" [9], seeing that the MSE is also highly correlated with the index in this regard. Probably the biggest challenge to the SSIM occurs in the event of a structural change. Linear correlation, as Wang et al. appropriately identify as "the cosine of the angle between the vectors $\mathbf{x} - \mu_x$ and $\mathbf{y} - \mu_y$" [9], seems to offer no meaningful advantage over the simple numerical errors determined by the mean squared error. More formally, this establishes a tangible relationship between the angle between two vectors, as given by the SSIM, and the average distance separating them, as calculated by the MSE. Together, the results of this study further weaken any credibility that the structural metric might have as a perceptual quality algorithm.

B. Sheikh et al.

Ironically, some of the strongest evidence against the structural algorithm comes from the developers of the metric themselves. In a recent investigation of a number of full reference metrics, the team, now headed by Sheikh, finds that the SSIM is linearly correlated with human-derived quality scores to a degree of 0.9393. The PSNR, and therefore the MSE, is itself able to earn a respectable score of 0.8709. One would surely expect a larger gap between the two in this respect, given the supposed advantages of the SSIM. The group nonetheless claims to be able to statistically differentiate between the two, at least at the 0.05 *level of confidence* [30]. Unfortunately, "the selection of the confidence criterion is also somewhat arbitrary and it obviously affects the conclusions being drawn" [33]. Even more suspicious is the discovery that "none of the IQMs evaluated in this study was statistically at par with the Null model on any of the datasets using a 95% criterion, suggesting that more needs to be done in reducing the gap between machine and human evaluation of image quality" [33]. The "Null model", more thoroughly described in [33], is said to be the event in which "residuals from one IQM come from the same distribution and are statistically indistinguishable (with 95% confidence) from the residuals from another IQM" [33]. This statement would seem to imply that even the designers themselves are unable to find any substantial divisions between the SSIM and the other techniques.

VII. Conclusion

In both an empirical study and a formal analysis, evidence of a systematic relationship between the structural similarity index and the dubious mean squared error is uncovered. While the statistical findings are convincing, the mathematical relations are undeniable. With these analytical formulas in hand, it is possible to derive the quality score of one metric given the other. Visual examples, namely equal-measure coded images and error maps, along with the conclusions of various external groups, bring the two techniques even closer. Though noble in their quest, Wang et al. have once again illustrated the enormous gap that continues to exist between a numerical measure of quality and that of the human mind. Until a more radical approach is taken, this problem will likely continue to confound researchers in the field.

Acknowledgment

This research was generously sponsored by the Natural Sciences and Engineering Research Council of Canada (NSERC). Additional resources were made available by the New Media Studio Lab (NMSL) at the University of Regina. Further thanks are extended to Dr. Shaun Fallat of the Department of Mathematics and Statistics at the University of Regina for his assistance in developing the mathematical foundations of this research.

References

[1] S. Winkler, *Digital Video Quality: Vision Models and Metrics*. West Sussex, England: John Wiley & Sons, Ltd., 2005.
[2] S. Süsstrunk and S. Winkler, "Color image quality on the internet," *Proc. SPIE Electronic Imaging: Internet Imaging*, vol. 5304, pp. 118–131, 2004.
[3] M. Miyahara, K. Kotani, and V. R. Algazi, "Objective picture quality scale (pqs) for image coding," *IEEE Trans. Communications*, vol. 46, pp. 1215–1226, September 1998.
[4] J. Lubin, *Vision Models for Target Detection and Recognition*. Singapore: World Scientific, 1995, ch. A Visual Discrimination Model for Imaging System Design and Evaluation, pp. 245–283.
[5] S. Daly, *Digital Images and Human Vision*. Cambridge, MA: MIT Press, 1993, ch. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, pp. 179–206.
[6] Z. Wang and A. C. Bovik, "Why is image quality assessment so difficult?" *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 4, pp. 3313–3316, May 2002.
[7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *Proc. IEEE Asilomar Conf. Signals, Systems and Computers*, vol. 2, pp. 1398–1402, November 2003.
[8] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, March 2002.
[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.
[10] Z. Wang, A. C. Bovik, and E. P. Simoncelli, *Handbook of Image and Video Processing*, 2nd ed. Academic Press, 2005, ch. 8.3: Structural Approaches to Image Quality Assessment, pp. 961–974.
[11] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. II, pp. 573–576, March 2005.
[12] L. Zhang, Z. Guo, Z. Wang, and D. Zhang, "Palmprint verification using complex wavelet transform," *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. 417–420, September 2007.
[13] S. Gupta, M. P. Sampat, M. K. Markey, and A. C. Bovik, "Facial range image matching using the complex wavelet structural similarity metric," *Proc. IEEE Workshop on Applications of Computer Vision*, p. 4, February 2007.
[14] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, February 2004.
[15] Z. Wang, H. R. Sheikh, and A. C. Bovik, *The Handbook of Video Databases: Design and Applications*. CRC Press, September 2003, ch. 41: Objective Video Quality Assessment, pp. 1041–1078.
[16] D. Chih-Che Lin and P. M. Chau, "Objective human visual system based video quality assessment metric for low bit-rate video communication systems," *Proc. IEEE Workshop on Multimedia Signal Processing*, pp. 320–323, October 2006.
[17] B. Coskun and B. Sankur, "Robust video hash extraction," *Proc. IEEE Signal Processing and Communications Applications Conf.*, pp. 292–295, April 2004.
[18] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R. W. Heath Jr., "Ssim-optimal linear image restoration," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 765–768, April 2008.

[19] M. P. Sampat, Z. Wang, M. K. Markey, G. J. Whitman, T. W. Stephens, and A. C. Bovik, "Measuring intra- and inter-observer agreement in identifying and localizing structures in medical images," *Proc. IEEE Int. Conf. Image Processing*, pp. 81–84, October 2006.
[20] Z. Wang, Q. Li, and X. Shang, "Perceptual image coding based on a maximum of minimal structural similarity criterion," *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. 121–124, September 2007.
[21] C.-C. Sung, S.-J. Ruan, B.-Y. Lin, and M.-C. Shie, "Quality and power efficient architecture for the discrete cosine transform," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E88-A, no. 12, pp. 3500–3507, December 2005.
[22] Z.-Y. Mai, C.-L. Yang, K.-Z. Kuang, and L.-M. Po, "A novel motion estimation method based on structural similarity for h.264 inter prediction," *Proc. IEEE. Int. Conf. Acoustics, Speech and Signal Processing*, vol. 2, pp. 913–916, May 2006.
[23] D. M. Rouse and S. S. Hemami, "Analyzing the role of visual structure in the recognition of natural image content with multi-scale ssim," *Proc. Western New York Image Processing Workshop*, October 2007.
[24] W. P. Cockshott, S. L. Balasuriya, I. P. Gunawan, and J. P. Siebert, "Image enhancement using vector quantisation based interpolation," *Proc. British Machine Vision Conf.*, September 2007.
[25] İ. Avcıbaş, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *Journal of Electronic Imaging*, vol. 11, no. 2, pp. 206–223, April 2002.
[26] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "A multidimensional image quality measure using singular value decomposition," *Proc. SPIE Image Quality and System Performance Conf.*, vol. 5294, pp. 82–92, January 2004.
[27] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Addison-Wesley, 1992.
[28] K. Seshadrinathan and A. C. Bovik, "New vistas in image and video quality assessment," *Proc. SPIE Human Vision and Electronic Imaging*, vol. 6492, January 2007.
[29] S. S. Vorren, "Subjective quality evaluation of the effect of packet loss in high-definition video," Norwegian University of Science and Technology Department of Electronics and Telecommunications, June 2006.
[30] M. F. Triola, *Elementary Statistics*. Pearson, 2005.
[31] A. Beghdadi and R. Iordache, "Image quality assessment using the joint spatial/spatial-frequency representation," *EURASIP Journal on Applied Signal Processing*, 2006.
[32] J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman, *Applied Linear Statistical Models*. Irwin, 1996.
[33] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440–3451, November 2006.