SIViP (2011) 5:81–91
DOI 10.1007/s11760-009-0144-1

ORIGINAL PAPER

A comprehensive assessment of the structural similarity index

Richard Dosselmann · Xue Dong Yang

Received: 29 June 2009 / Revised: 15 October 2009 / Accepted: 16 October 2009 / Published online: 6 November 2009

Abstract In recent years the structural similarity index has become an accepted standard among image quality metrics. Made up of three components, this technique assesses the visual impact of changes in image luminance, contrast, and structure. Applications of the index include image enhancement, video quality monitoring, and image 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 often unreliable, mean squared error. In the first evaluation, the two metrics are statistically compared with one another. Then, in the second, a pair of functions that algebraically connects the two is derived. These results suggest a much closer relationship between the structural similarity index and mean squared error.

Keywords Structural similarity index · SSIM · Mean squared error · MSE · Image quality metric

1 Introduction

Image quality assessment is an emerging area in signal processing. More or less defined as the task of designing an algorithm to automatically judge the perceived visual "quality" of an image, it remains a largely open problem. While a variety of algorithms, or *image quality metrics* (IQM) [26,39], have been proposed, none truly correlates with the notion of "quality" as perceived by the *human visual system* (HVS) [39]. Though several attempts [7,14,16] have been made to simulate the intricate processes of the HVS, most approaches generally describe quality in terms of the numerical pixel differences between an *original* image and its corrupted, or *coded*, counterpart. For a given image, its original form is one that is free of any distortions and is therefore assumed to be of perfect quality. Applications of image quality include judging the performance of compression algorithms, television and video monitoring [9], and automatic image enhancement.

A short time ago, a new approach based on statistical changes in image *luminance* [24], *contrast* [24] and *structure* [32], was put forward. Known as the *structural similarity* (SSIM) index [32,33,35,38], this method has quickly become the subject of considerable research and attention. The product of several years of research [29,30,32,33,38] itself, this metric has been formulated and revised a number of times. It has been employed in image restoration [3], video quality monitoring [4,6,35,36], image enhancement [5], video compression [15,25], visual cognition [18], and imaging coding [34]. Further applications are described in [31]. Despite the growing number of applications, this research, prompted by the popularity of the metric, finds that the SSIM is both statistically and analytically linked to perhaps the simplest and oldest of all IQMs, namely the *mean squared error* (MSE) [39]. Hence, the objective of this paper is to establish a relationship between the SSIM and MSE. This work is not the first to question the SSIM. Various external studies, such as those of [2,17,28], also raise concerns. This research, however, appears to be the first to directly consider the statistical relationships between the two methods. As well, this work develops a pair of mathematical functions

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.

R. Dosselmann (✉) · X. D. Yang
Department of Computer Science, University of Regina,
3737 Wascana Parkway, Regina, SK S4S 0A2, Canada
e-mail: dosselmr@cs.uregina.ca

X. D. Yang
e-mail: yang@cs.uregina.ca

that directly link the two. Given these findings, one is left to question whether the structural similarity index is ready for widespread adoption.

The two metrics considered in this investigation are formally introduced and discussed in Sects. 2 and 3, respectively. The SSIM is then thoroughly evaluated in Sects. 4 and 5, before closing statements are made in Sect. 6.

2 Mean squared error

Perhaps no metric has received more attention than the mean squared error. Its simple formulation and clear interpretation have allowed this technique to become one of the most widely used metrics in the field. Unfortunately, it often produces misleading values that do not correlate well with perceived quality [23,26,39]. This is not surprising given that the metric is nothing more than a measure of the per-pixel differences between an original image and its distorted form.

2.1 Definition

Formally, the mean squared error is defined as

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

where n is the number of pixels in an n-dimensional image vector, and x_i and y_i denote the gray levels of 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 $\mathbf{x}$ and $\mathbf{y}$. An MSE score is given on the interval $[0, \infty)$. A low value indicates a small error and therefore a high level of quality. Conversely, a large score is associated with a more significant error and, accordingly, lower visual quality.

2.2 Equal MSE hypersphere

One of the most frequently cited problems of the MSE is its inability to differentiate perceived quality across different distortions. For instance, although an image in which there has been a shift in luminance is normally of higher visual quality than one that is tainted by compression, the MSE, being highly susceptible to large numerical changes in gray levels, improperly concludes that the former image is of lower quality than the latter. Unfortunately, this problem extends far beyond luminance and compression. In fact, for a variety of distortions, it is possible to create a series of distorted images that are nearly identical in terms of their MSE, yet are visually distinct in terms of their perceived quality. Arranging these images, equidistant from a single original image, one constructs an *equal MSE hypersphere* [33].

This concept is readily illustrated in the example of Fig. 1. Here, each of the images is approximately equidistant from the original, despite their striking differences in perceived quality. The first of the two noisy images, seen in Fig. 1b, appears to have only been marginally affected by the superimposed noise. Though somewhat more noticeably corrupted, the second noisy image, given in Fig. 1c, is also at a similar distance from the original. Conversely, the blurred and compressed images of Fig. 1d, e and f are much worse, though the MSE suggests that they are comparable to Fig. 1b and c. Clearly, the performance of the MSE is highly suspect.

3 Structural similarity index

Given the obvious limitations of the mean squared error, Wang et al. [32,33,35,38] propose the structural similarity index as a more involved solution to the problem of image quality assessment. Made up of three terms, the index estimates the visual impact of shifts in image luminance, changes in 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" [38].

According to its designers, the "SSIM correlates extraordinarily well with perceptual image quality, and handily outperforms prior state-of-the-art HVS-based metrics" [20]. 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" [20]. When compared with the mean squared error, "the SSIM index has been shown to outperform MSE and the related PSNR[1] in measuring the quality of natural images across a wide variety of distortions" [3]. In a direct contradiction to this statement, Reibman finds that the "MSE outperforms SSIM for all pooling strategies" [17]. Additionally, Vorren "cannot statistically prove that SSIM is better, due to the overlapping of the confidence intervals of all the video quality models" [28]. Some of the strongest contradicting evidence comes from the developers of the SSIM metric themselves. First, the graphical plots of Figs. 3 of [38], 5 of [35], and 7 of [32] suggest that the PSNR is almost as well correlated with human-derived subjective quality scores as is the SSIM. Second, in a recent investigation [22] of a number of quality metrics, the team finds the SSIM to be *correlated* [27] 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

[1] The PSNR, or *peak signal-to-noise ratio*, is an adjusted form of the MSE. It is formally defined in [39].

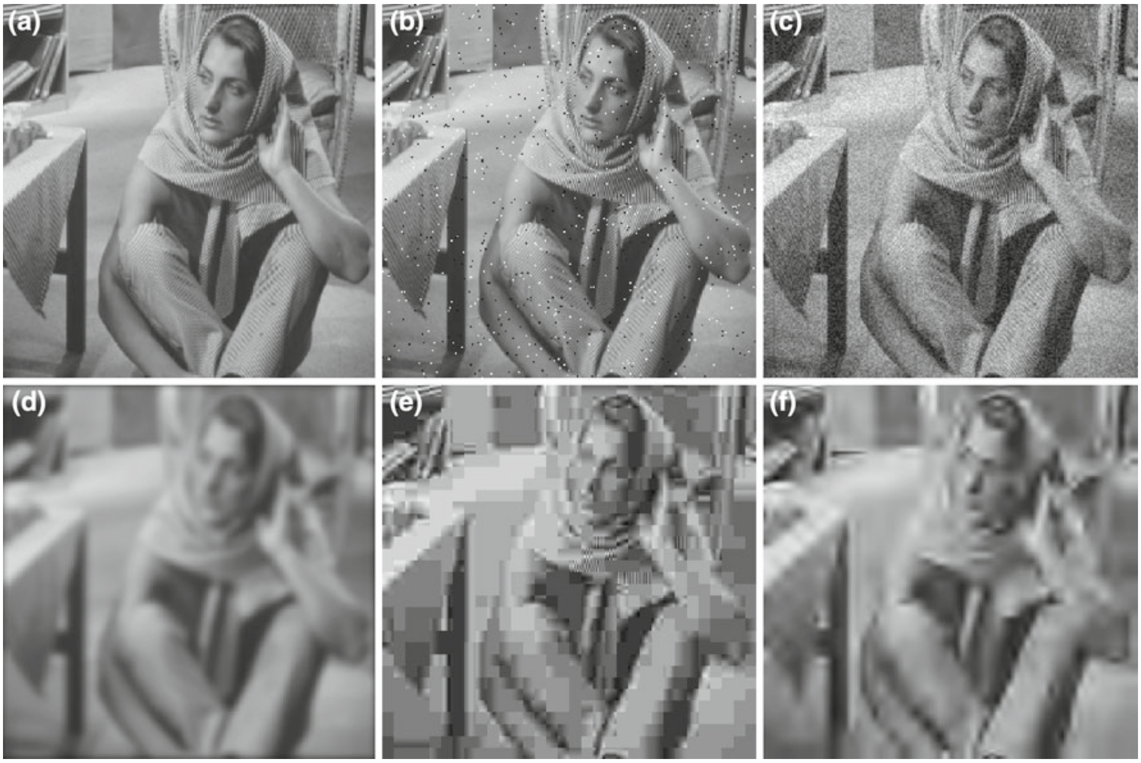

Fig. 1 Equal MSE hypersphere: **a** Original “Barbara” image; **b** 1.4% salt and pepper noise, MSE = 259.9516; **c** 11.35σ Gaussian noise, MSE = 258.6884; **d** 11 × 11 Gaussian blur, MSE = 257.7267; **e** 90% JPEG compression, MSE = 259.6023; **f** 98% JPEG2000 compression, MSE = 270.2315

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* [27]. The group admits, however, that “the selection of the confidence criterion is also somewhat arbitrary and it obviously affects the conclusions being drawn” [22]. Moreover, the team concludes that “none of the IQMs evaluated in this study was statistically at par with the Null model[2] 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” [22]. In a more recent paper, the group admits that in fact “the structure term in the SSIM index computes an MSE between normalized image patches” [21].

3.1 Definition

For original and coded images **x** and **y**, respectively, the SSIM index is defined as

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

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. [32], $\alpha = \beta = \gamma = 1$. The luminance, contrast, and structural components of the index are defined individually as

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

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

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

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, respectively, σ_x^2 and σ_y^2 denote the variances, respectively, and σ_{xy} is the covariance of the two images. 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 [32] are introduced. For an 8-bit grayscale image with a maximum gray level of $L = 255$, $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$ and $C_3 = C_2/2$, where $K_1 = 0.01$ and $K_2 = 0.03$ [35]. When $C_1 = C_2 = 0$, the metric is reduced to the *universal quality index* (UQI) [29], a predecessor of the SSIM index. Note that (5), absent C_3, represents the *linear correlation* [27] of the two images.

[2] The “Null model” is described in [22].

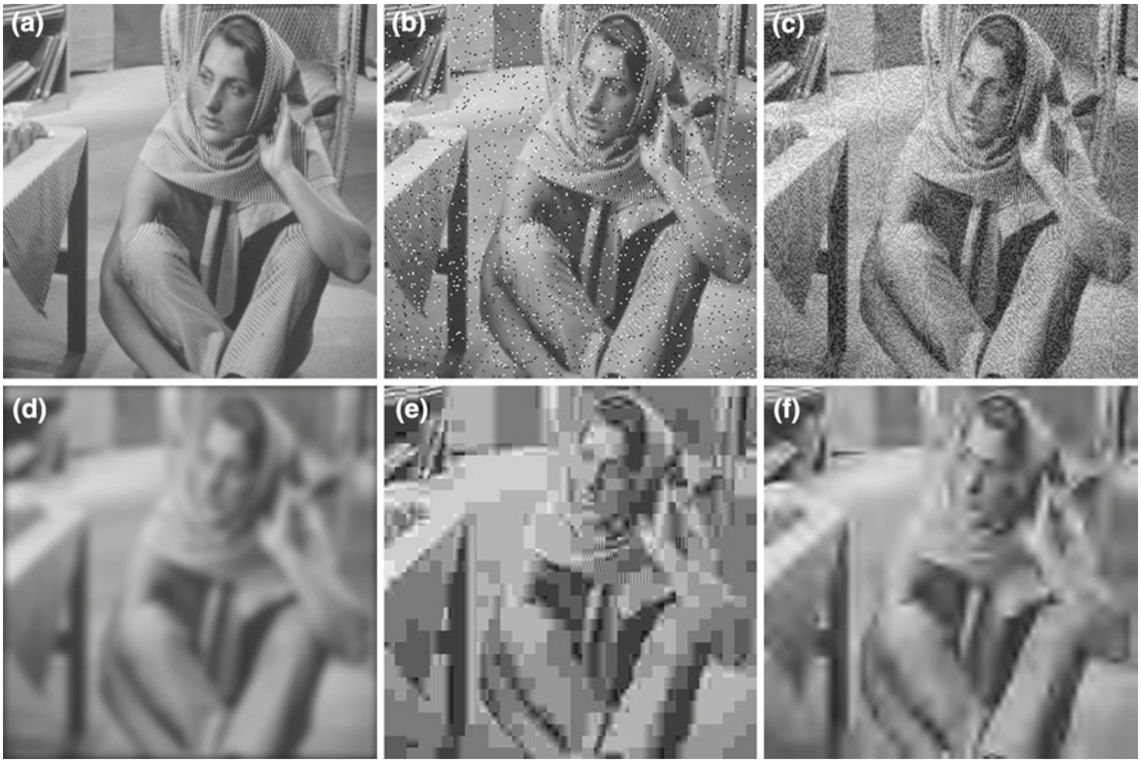

Fig. 2 Equal SSIM hypersphere: **a** original "Barbara" image; **b** 4% salt and pepper noise, SSIM = 0.4950; **c** 16.00σ Gaussian noise, SSIM = 0.4958; **d** 13 × 13 Gaussian blur, SSIM = 0.4886; **e** 95% JPEG compression, SSIM = 0.4531; **f** 98% JPEG2000 compression, SSIM = 0.4980

The formulation given in (2) satisfies three properties. First, it ensures *symmetry* [32], meaning that $\text{SSIM}(\mathbf{x}, \mathbf{y}) = \text{SSIM}(\mathbf{y}, \mathbf{x})$. Thus, the original and coded images may be swapped. As well, the metric guarantees *boundness* [32], in the sense that $-1 \leq \text{SSIM}(\mathbf{x}, \mathbf{y}) \leq 1$. In most cases, however, a score is given on the interval [0, 1], where values closer to 0 represent lower levels of image quality while values nearer to 1 are indicative of higher levels of visual quality. Finally, there is a *unique maximum* [32], meaning that $\text{SSIM}(\mathbf{x}, \mathbf{y}) = 1$ if and only if $\mathbf{x} = \mathbf{y}$.

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 image, referred to as the *mean SSIM* (MSSIM) [32],[3] is the arithmetic average of the local scores.

3.2 Equal SSIM hypersphere

Wang et al. often highlight the dubious behavior of the MSE by way of an equal MSE hypersphere. Consider the *equal SSIM hypersphere* of Fig. 2. Like the equal MSE hypersphere of Sect. 2.2, each of the coded images in this example is at an almost equal distance from the original, this time in terms of its SSIM score. And, just as the MSE does, the SSIM erroneously concludes that the noisy images of Fig. 2b and c are visually equivalent to the blurred and compressed images of Fig. 2d, e and f. Note that this research is not the first to unearth such problems. The index also incorrectly orders images in Figs. 3 and 4 of [2].

In each of Figs. 1 of [37], 2 of [29], 2 of [32], 2 of [31], 3 of [30], 4 of [33] and 41.9 of [36], Wang et al. showcase an example of a hypersphere that is quite different from that of Fig. 2. This example, replicated in Fig. 3, is one in which the SSIM correctly orders the images, yet the MSE is unable to detect any measurable differences. This example seems to be a bit unique. First, by adjusting luminance and contrast, one obtains the images given in Fig. 3b and c. Though the SSIM is not significantly impacted by these changes, given that it is equipped to address such issues, the MSE responds by assigning high error scores. From here, one is then able to carefully generate three more images with equally large MSE scores. In the first of these three, namely Fig. 3d, a mere dash of noise is added, presumably since the SSIM, like the MSE, seems to be rather sensitive to noise. A more destructive blur is then introduced to produce the coded image of Fig. 3e. Last, Fig. 3f is heavily compressed with JPEG, producing

[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 windows.

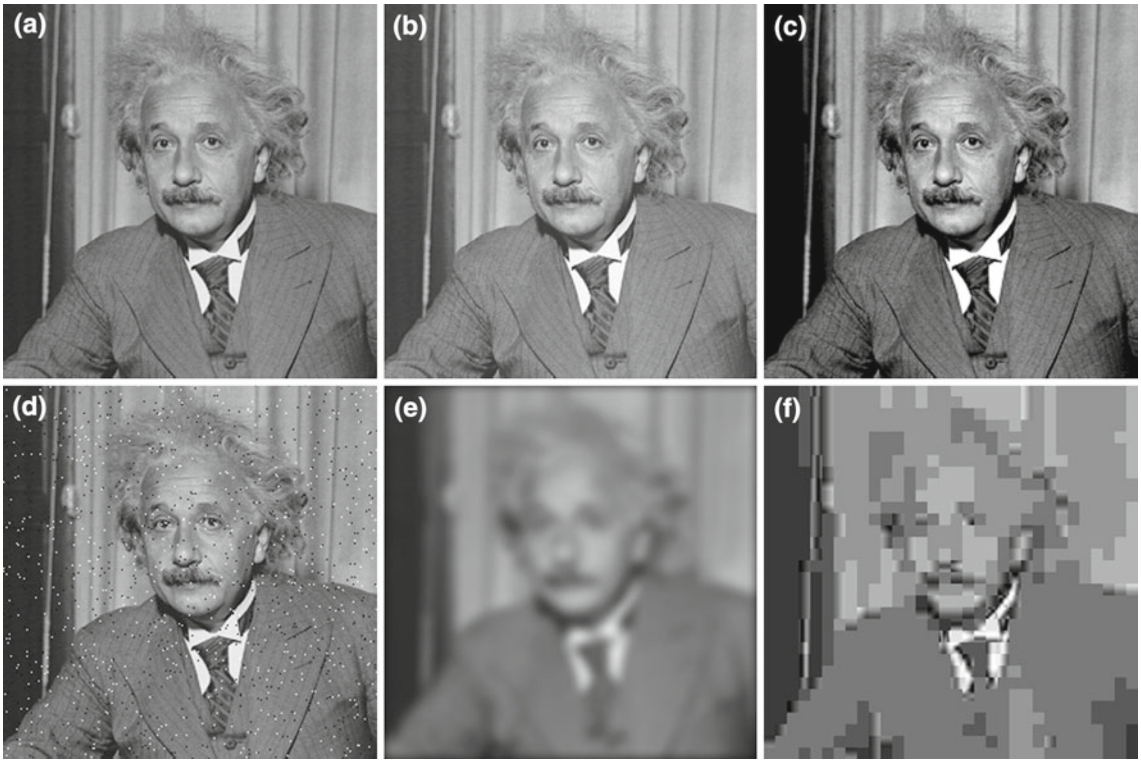

Fig. 3 Alternate equal MSE hypersphere: **a** original "Einstein" image; **b** 7% luminance shift, SSIM = 0.9849, MSE = 323.8797; **c** 33% contrast adjustment, SSIM = 0.8876, MSE = 328.1693; **d** 1.8% salt and pepper noise, SSIM = 0.6410, MSE = 322.2579; **e** 15 × 15 Gaussian blur, SSIM = 0.3040, MSE = 343.9443; **f** 99% JPEG compression, SSIM = 0.2736, MSE = 323.7193

an image that is highly corrupted, both visually and numerically. The introduction of additional noise, for instance, or a decrease in the amount of the blur will typically result in incorrect scores on the part of the SSIM. Specifically, added noise, to which the SSIM is fairly sensitive, abruptly brings down the overall quality score of Fig. 3d. Consequently, it is improperly assigned a lower quality score than the blurred and compressed images of Fig. 3e and f. In an opposite fashion, decreasing the amount of blur or compression in Fig. 3e or f causes the scores of these images to erroneously rise above that of Fig. 3d. It is for reasons such as these that this particular hypersphere appears to be rather special. In a more typical situation, such as that of Fig. 2, in which changes in luminance and contrast are absent, the SSIM performs just as poorly as the MSE.

4 Empirical assessment

In this statistical evaluation[4] of the SSIM index, the scoring behavior of the metric is compared with that of the MSE. Ensuing statistical analysis suggests a high level of association between the two.

4.1 Method

As part of the experiment, a set of 15 "classic" test images, shown in Fig. 4, was assembled. They vary widely in terms of shading, texture, level of detail, and content. To illustrate the procedure used in the experiment, consider, as an example, the image of Fig. 4a ("Barbara"). This image is synthetically corrupted using a variety of distortions to produce ten individual coded images. The exact type and amount of distortion used to generate each of these ten images is more thoroughly described in the following. Next, each of these ten coded images is compared with the original of Fig. 4a using both the SSIM and MSE metrics, yielding two sets of quality scores. Within each of these two sets, the ten coded images are then assigned *ranks* [27] based on their quality scores, with the image of lowest measured quality receiving a rank of 1 and the best being given a rank of 10. Last, the degree of correlation between the ranks in the two sets is measured using *Spearman's rank correlation coefficient* [27], denoted by r. Much like linear correlation, an r value is given on the interval $[-1, 1]$, where a score near 0 indicates

[4] The empirical assessment presented in this paper is separate from that of [10].

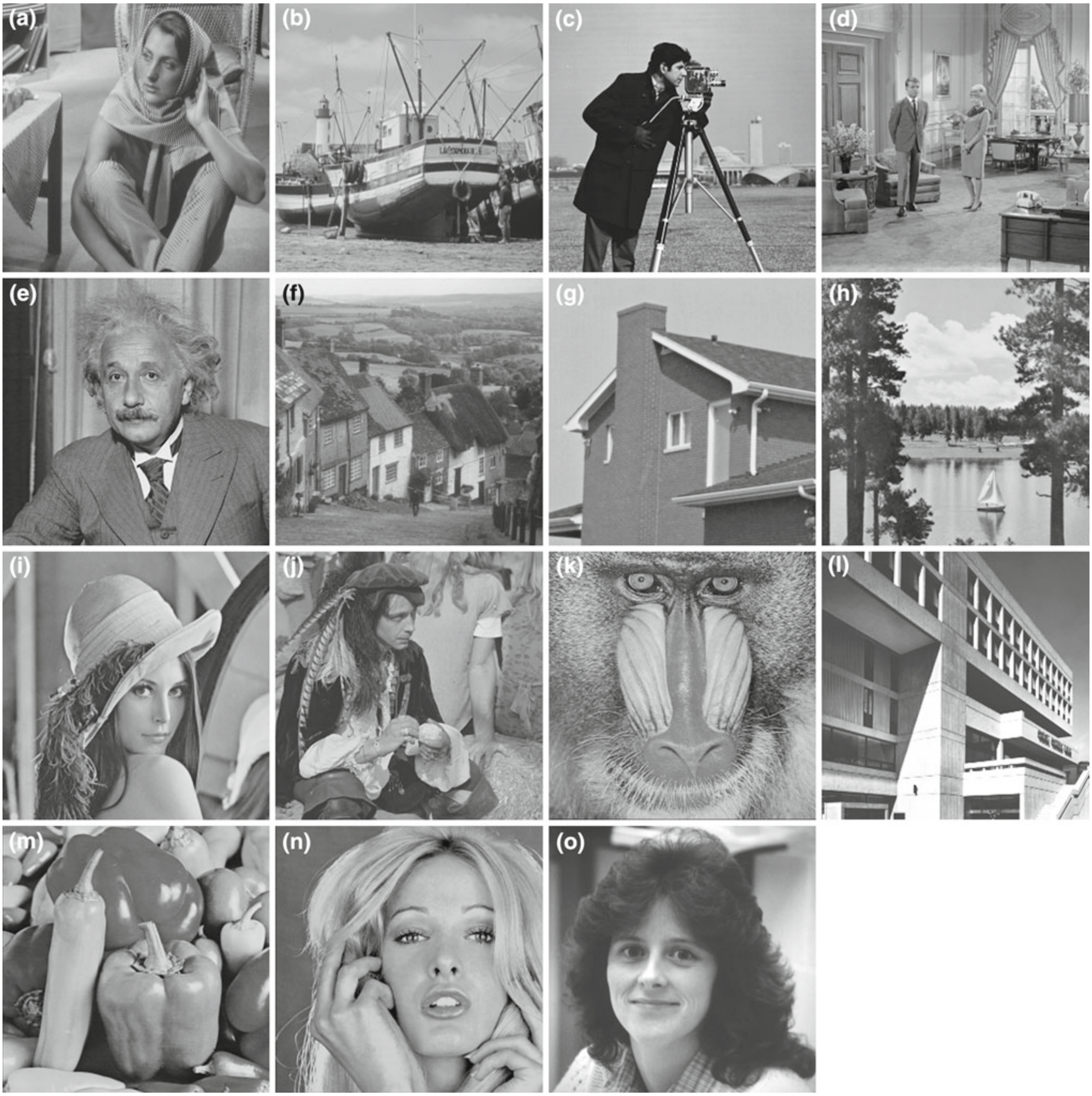

Fig. 4 Test images: **a** "Barbara"; **b** "boat"; **c** "cameraman"; **d** "couple"; **e** "Einstein"; **f** "Goldhill"; **g** "house"; **h** "lake"; **i** "Lena"; **j** "man"; **k** "mandrill"; **l** "MIT"; **m** "peppers"; **n** "Tiffany"; **o** "woman"

little association between two variables, while a value closer to ±1 is evidence of a stronger degree of correspondence. Hence, this experiment seeks to confirm that both the SSIM and MSE order distorted images in the same manner, just as the examples of Figs. 1 and 2 suggest. This entire process is repeated five times for each of the 15 test images of Fig. 4. Thus, for each of the 15 test images, there are five unique sets, labeled as S_1 through S_5, of ten coded images. Each such set is assigned a separate correlation coefficient based on its rank orderings. In total, for all 15 test images, there are $15 \times 5 = 75$ sets and therefore 75 correlation coefficients.

The precise impairments considered in this experiment consist of the addition of *salt and pepper* [11], *Gaussian* [11] and *sinusoidal* or *periodic* [11] noise, blurring using *averaging filters* [11] and *Gaussian smoothing* [13], *JPEG*[5]

[5] JPEG- and JPEG2000-compressed images were generated using Jasc® Paint Shop Pro.

Table 1 MSE versus SSIM Spearman rank correlation coefficients (r)

Image	S_1	S_2	S_3	S_4	S_5
"Barbara"	0.9394	0.9394	0.9636	0.9758	0.9758
"boat"	1.0000	0.9758	0.9030	0.8909	0.7212
"cameraman"	0.9515	0.9030	0.8303	0.9636	1.0000
"couple"	0.8061	0.9394	0.8303	0.9030	0.8061
"Einstein"	0.9879	0.8182	0.9636	0.9152	0.8667
"Goldhill"	0.6727	0.9152	0.7818	0.9636	0.9879
"house"	0.8667	0.7333	0.9636	0.8788	0.9515
"lake"	0.6970	0.9515	0.8667	0.9515	0.8182
"Lena"	0.9152	1.0000	0.9879	0.9030	0.9636
"man"	0.9152	0.6364	0.9636	0.9030	0.9273
"mandrill"	0.9515	0.7212	0.8182	0.8909	0.9152
"MIT"	0.9273	0.9394	0.9273	0.9879	0.9394
"peppers"	0.9030	0.9879	0.9152	0.9879	0.9273
"Tiffany"	0.9879	1.0000	0.9879	0.9636	0.9394
"woman"	0.9879	0.9636	0.9394	0.9879	0.9879

Table 2 MSE and SSIM ranks for set S_5 of "mandrill" image

Coded image	MSE	SSIM	MSE rank	SSIM rank
I_1	78.0649	0.9428	7	7
I_2	395.9164	0.3555	5	3
I_3	567.3937	0.1175	3	1
I_4	9.1176	0.9704	9	9
I_5	230.5747	0.6695	6	6
I_6	27.2760	0.9495	8	8
I_7	14653.3166	0.1477	1	2
I_8	3.5099	0.9958	10	10
I_9	691.5599	0.5090	2	4
I_{10}	478.0456	0.5795	4	5

I_1: 4 lowest-order bits truncated; I_2: 7 × 7 Gaussian blur; I_3: 15 × 15 Gaussian blur; I_4: 55% JPEG2000 compression; I_5: 3 × 3 Gaussian blur; I_6: 15% JPEG compression; I_7: 6 Laplacian sharpenings; I_8: 2 lowest-order bits truncated; I_9: 18.60σ Gaussian noise; I_{10}: 25 sinusoidal noise impulses

[11] and *JPEG2000* [1] compression, *Laplacian* [11] sharpening and *quantization* [11]. Salt and pepper noise is injected in random quantities, ranging from 1.0 to 25.0%. Gaussian noise is produced by randomly choosing a new gray level that is within a specified number of standard deviations of the current gray level. This standard deviation, represented by σ, is anywhere from 1.00 to 25.00. Periodic noise is implemented by arbitrarily introducing between 1 and 25 impulses into the frequency domain of a given image. In the case of blurring, the sizes of the averaging and Gaussian filters are randomly selected, with exact sizes varying between 3 × 3 and 49 × 49. Both JPEG and JPEG2000 are applied in amounts ranging from 5 to 99%. The degree of Laplacian sharpening is determined by the number of times that the sharpening filter is applied. In particular, sharpening is performed anywhere from 1 to 10 times. Finally, errors in quantization are simulated by randomly setting either 1, 2, 3 or all 4 of the lowest-order bits of each pixel in an image to zero. Although the aforementioned errors appear along side of one another in the various corrupted images in each of the 75 data sets, at no point do any two occur together in a single image. As an example, a set may contain an image tainted by 15% Gaussian noise and another warped by a 5 × 5 averaging blur, but not an image in which both of these errors simultaneously occur.

4.2 Results and discussion

Individual r correlations for each of the 75 test sets are found in Table 1. These findings suggest a reasonably significant level of correlation between the SSIM and MSE. Values range from $r = 0.6364$ to $r = 1.0000$, with an average of $r = 0.9116$ and a variance of 0.007. An average this large, along with a small variance, suggests that most of the correlations are decidedly significant. Clearly, when ordering coded images, the SSIM and MSE often choose similar arrangements. Results such as this are likely a sign of a deeper relationship between the two methods.

As an example, the ranks assigned to the ten coded images of set S_5 of Fig. 4k ("mandrill") are shown in Table 2. This example was chosen because of the closeness of its coefficient, $r = 0.9152$, to the average. The allotted ranks of Table 2 are evidence of a general trend between the metrics. While the mean squared error is more responsive to noise, the structural similarity index is more sensitive to blurring. Because it is measured locally, the SSIM is understandably less reactive to noise, but more so to blurring. There are no disparities in the instances of compression and quantization.

5 Formal assessment

The results of the preceding sections suggest comparable performance on the part of the structural similarity index and mean squared error. Not surprisingly, one is therefore motivated to explore the possibility that the two may be analytically related. In the ensuing section, the two are algebraically shown to be functions of one another, thereby further underscoring their connection.

5.1 Method

First, the definition of the SSIM given in (2) is expanded and reorganized. In order to simplify an otherwise lengthy computation, the constants C_1, C_2 and C_3 of (2) are omitted. Several terms are then added to obtain an expression equal to the MSE. This first expression is subsequently reordered to obtain the second mapping function. As an added remark, Rouse and Hemami [19] also propose a reorganized form of the SSIM, albeit distinct from that of this work.

For a $k = 64$-pixel 8×8 window, the simplified SSIM, without the constants C_1, C_2 and C_3, is given as

$$\begin{aligned}
&\mathrm{SSIM}^*(\mathbf{x},\mathbf{y}) \\
&= \left(\frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2}\right)\left(\frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)\left(\frac{\sigma_{xy}}{\sigma_x\sigma_y}\right) && (6)\\
&= \frac{4\mu_x\mu_y\sigma_{xy}}{\left(\mu_x^2+\mu_y^2\right)\left(\sigma_x^2+\sigma_y^2\right)} && (7)\\
&= \frac{4\mu_x\mu_y\left(\frac{1}{k}\sum_{i=1}^{k}x_iy_i-\mu_x\mu_y\right)}{\left(\mu_x^2+\mu_y^2\right)\left[\left(\frac{1}{k}\sum_{i=1}^{k}x_i^2+\frac{1}{k}\sum_{i=1}^{k}y_i^2\right)-\left(\mu_x^2+\mu_y^2\right)\right]} && (8)\\
&= \frac{k_2\left(\frac{2}{k}\sum_{i=1}^{k}x_iy_i-k_2\right)}{k_1\left(l_1+l_2-k_1\right)}, && (9)
\end{aligned}$$

where $l_1 = \frac{1}{k}\sum_{i=1}^{k}x_i^2$, $l_2 = \frac{1}{k}\sum_{i=1}^{k}y_i^2$, $k_1 = \mu_x^2+\mu_y^2$, and $k_2 = 2\mu_x\mu_y$. Cross multiplication in (9) gives

$$k_1\left(l_1+l_2-k_1\right)\mathrm{SSIM}^*(\mathbf{x},\mathbf{y}) = k_2\left(\frac{2}{k}\sum_{i=1}^{k}x_iy_i-k_2\right) \tag{10}$$

$$\Rightarrow -\frac{k_1\left(l_1+l_2-k_1\right)\mathrm{SSIM}^*(\mathbf{x},\mathbf{y})}{k_2}-k_2 = -\frac{2}{k}\sum_{i=1}^{k}x_iy_i \tag{11}$$

$$\begin{aligned}
\Rightarrow &-\frac{k_1\left(l_1+l_2-k_1\right)\mathrm{SSIM}^*(\mathbf{x},\mathbf{y})}{k_2}\\
&+l_1+l_2-k_2 = \mathrm{MSE}^*,
\end{aligned} \tag{12}$$

where MSE^* is a local version of the mean squared error calculated over an 8×8 window. Rearranging the elements of (12), the equation is solved for $\mathrm{SSIM}^*(\mathbf{x},\mathbf{y})$. Together, the two mapping functions are given as

$$\mathrm{MSE}^* = -\frac{k_1\left(l_1+l_2-k_1\right)\mathrm{SSIM}^*(\mathbf{x},\mathbf{y})}{k_2}+l_1+l_2-k_2 \tag{13}$$

and

$$\mathrm{SSIM}^*(\mathbf{x},\mathbf{y}) = -\frac{k_2\left(\mathrm{MSE}^*-l_1-l_2+k_2\right)}{k_1\left(l_1+l_2-k_1\right)}. \tag{14}$$

Equations (13) and (14) make use of a local version of the mean squared error, namely MSE^*, despite the fact that the metric is actually measured at the global scope. Fortunately, with a few modifications, a local MSE^* value is readily transformed into a global measure. To do so, one first determines the MSE^* for each 8×8 window. These local scores are then added to produce a global sum, denoted by S. Next, notice that as the sliding 8×8 window moves across the image, each pixel in the interior portion of an image is processed exactly $8 \times 8 = 64$ times. Pixels lying along the edges of the image, however, are scanned fewer times. For example, the four pixels located at the corners of the image each pass under the sliding 8×8 window only once and are therefore processed only one time each. To compensate, these missing terms must be added to S. For each pair of edge pixels x_i and y_i in the original and coded images, respectively, each of which is processed $1 \le c_i < 64$ times, the value

$$\left(\frac{64-c_i}{64}\right)\left(x_i-y_i\right)^2 \tag{15}$$

is added to S. Following this adjustment, S is divided by the size of the image, namely n, yielding a value equal to that of the actual MSE.

Pseudocode of the algorithms to convert between the two techniques is given below. Here, W represents the total number of 8×8 windows in an image and $f_b(\mathbf{x},\mathbf{y})$ is a function that amends the final total to include the missing pixels along the image boundary, as described in the previous paragraph. The first of the two procedures, given in Algorithm 1, switches an SSIM score to its matching MSE counterpart.

Algorithm 1 SSIM to MSE

```
MSE ← 0
for each 8 × 8 window do
    $\mathrm{SSIM}(\mathbf{x},\mathbf{y})$ ← (2)
    $\mathrm{MSE}^*$ ← (13) using $\mathrm{SSIM}(\mathbf{x},\mathbf{y})$
    MSE ← MSE + $\mathrm{MSE}^*$
end for
MSE ← MSE + $f_b(\mathbf{x},\mathbf{y})$
MSE ← $\frac{1}{n}$MSE
```

The reverse transformation is given in Algorithm 2.

Algorithm 2 MSE to SSIM

```
$\mathrm{SSIM}(\mathbf{x},\mathbf{y})$ ← 0
$W$ ← 0
for each 8 × 8 window do
    $\mathrm{MSE}^*$ ← (1) using $k$
    $\mathrm{SSIM}^*(\mathbf{x},\mathbf{y})$ ← (14) using $\mathrm{MSE}^*$
    $\mathrm{SSIM}(\mathbf{x},\mathbf{y})$ ← $\mathrm{SSIM}(\mathbf{x},\mathbf{y}) + \mathrm{SSIM}^*(\mathbf{x},\mathbf{y})$
    $W$ ← $W + 1$
end for
$\mathrm{SSIM}(\mathbf{x},\mathbf{y})$ ← $\frac{1}{W}\mathrm{SSIM}(\mathbf{x},\mathbf{y})$
```

5.2 Results and discussion

Initially, (13) and (14) might look somewhat involved. Closer inspection, though, reveals that all but one of the terms l_1, l_2, k_1, and k_2 are for the most part irrelevant. The first of these four, namely l_1, may be essentially ignored seeing that

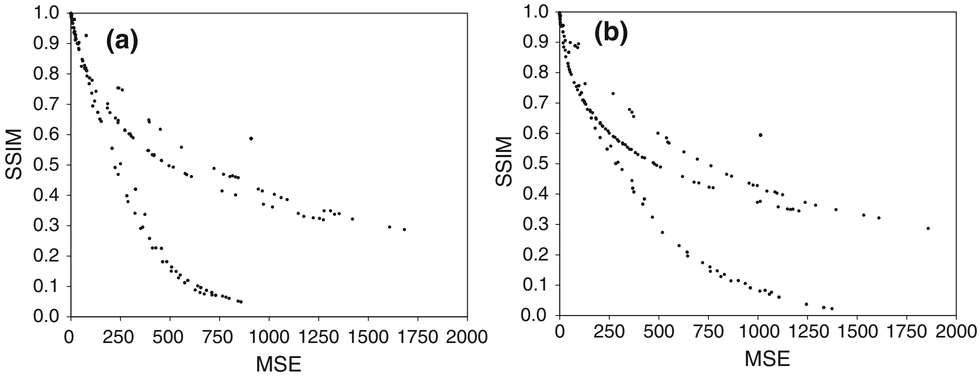

Fig. 5 Plots of MSE versus SSIM: **a** “Goldhill”; **b** “man”

it is nothing more than the sum of the squares of the gray levels of an original image and is unaffected by any errors. Additionally, experimental evidence indicates that both k_1 and k_2, functions of the means of the original and coded images, remain somewhat constant, with only minor fluctuations detected. This leaves only l_2, which represents the sum of the squares of the gray levels of the associated coded image. In general, as the amount of error in a coded image varies so does the value of this term. Thus, overlooking the relatively small l_1, k_1 and k_2 terms, the SSIM and MSE differ by only l_2, a simple proportional measure of error. This term does not indicate where such distortions occur or how perceptible they might be to a human observer. The MSE is more affected by large values of l_2 since this sum is added to the metric, as seen in (13). The addition of this factor helps to explain the slight variations in the rank orders of Sect. 4.2 that result from the introduction of noise. Beyond the l_2 term, the SSIM and MSE are both essentially measures of the *inner product* [12] of two images, as measured by the term

$$\sum_{i=1}^{k} x_i y_i \tag{16}$$

in each of the expansions of (9) and (11). As a side note, the presence of the l_2 term means that both (13) and (14) are multivariate functions, encompassing the three variables $\text{SSIM}^*(\mathbf{x}, \mathbf{y})$, MSE^* and l_2.

The relationship between the SSIM and MSE metrics is visualized in the two plots of Fig. 5a and b. In each of these two situations, 250 coded images were randomly created using the images of Fig. 4f (“Goldhill”) and j (“man”), respectively, and the same types of errors as those of Sect. 4. Comparing each of these coded images with its original, two sets of 250 SSIM and MSE scores are generated. Each coded image is then represented as a point in its respective plot, with its horizontal position determined by its MSE score and its vertical alignment given by its SSIM score. The multivariate nature of the relationship between the SSIM and MSE is immediately evident from the plots. Both plots are made up of a number of overlapping sequences of points. Each such sequence corresponds with a distinct type of error, such as noise or blurring. This occurrence is not at all unusual, given the unique effects of the various types of errors on the k_1, k_2 and l_2 terms, with most such variation arising in the l_2 term. For example, as blurring increases, the overall gray levels of a coded image are progressively reduced, causing the l_2 term to drop accordingly. Conversely, added noise rapidly inflates the value of this term. Compression, on the other hand, does little to change this value. These dissimilar behaviors give rise to the separate curves of Fig. 5a and b.

Using ten of the 250 coded images of Fig. 4j, the mapping functions of (13) and (14) are tested. The results are given in Table 3. By and large, results are fairly accurate. Numerical inaccuracies are attributed to the absence of the constants C_1, C_2 and C_3. Unfortunately, as commented earlier, the inclusion of these terms needlessly complicates the mapping functions. As these constants are used to ensure the numerical integrity of the SSIM metric and therefore have nothing to do with actual quality assessment, they may be safely removed. Either way, these results illustrate the functional dependency that exists between the two metrics.

6 Conclusion

In both an empirical study and a formal analysis, evidence of a relationship between the increasingly popular structural similarity index and the conventional mean squared error is uncovered. This research is perhaps the first to uncover a statistical link of this nature and likely the only in which a formal connection is established. Visual examples, namely the equal hypersphere images of Sects. 2.2 and 3.2, along with the discoveries of a number of external groups, bring the two techniques even closer together. Collectively, these findings suggest that the performance of the SSIM is perhaps much closer to that of the MSE than some might

Table 3 Computing MSE from SSIM and vice versa using "man" image

Coded image	Actual MSE	Actual SSIM	SSIM to MSE	MSE to SSIM
I_1	204.6116	0.5858	197.9315	0.5769
I_2	760.5304	0.1455	734.7558	0.1368
I_3	96.9155	0.7567	93.3684	0.7485
I_4	0.4212	0.9948	0.3996	0.9932
I_5	78.2636	0.8881	74.7704	0.8776
I_6	1012.9693	0.5943	977.8988	0.5930
I_7	1332.0771	0.0260	1297.0787	0.0150
I_8	119.1822	0.7099	113.9132	0.7086
I_9	95.5605	0.8948	92.001	0.8945
I_{10}	244.1873	0.6076	233.3428	0.6066

I_1: 80% JPEG compression; I_2: 21 × 21 Gaussian blur; I_3: 50% JPEG compression; I_4: 15% JPEG2000 compression; I_5: 4 lowest-order bits truncated; I_6: 1 Laplacian sharpening; I_7: 45 × 45 Gaussian blur; I_8: 7.70σ Gaussian noise; I_9: 1.4% salt and pepper noise; I_{10}: 17 sinusoidal noise impulses

claim. Consequently, one is left to question the legitimacy of many of the applications of the SSIM. Ultimately, this investigation once again illustrates the enormous gap that continues to exist between an automated measure of quality and that of the human mind. Until a more radical approach is considered, this problem will likely continue to confound researchers in the field. Ideas that explore this concept of a more "sophisticated" image quality metric are discussed in [8].

Acknowledgments The authors of this paper wish to thank 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. Acharya, T., Tsai, P.S.: JPEG2000 Standard for Image Compression: Concepts, Algorithms and VLSI Architectures. John Wiley & Sons, Inc., New York (2005)
2. Beghdadi, A., Iordache, R.: Image quality assessment using the joint spatial/spatial-frequency representation. EURASIP J. Appl. Signal Process. **2006**, 1–8 (2006)
3. Channappayya, S.S., Bovik, A.C., Caramanis, C., Heath Jr., R.W.: Ssim-optimal linear image restoration. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 765–768 (2008)
4. Chih-Che Lin, D., Chau, P.M.: Objective human visual system based video quality assessment metric for low bit-rate video communication systems. In: Proceedings of the IEEE Workshop on Multimedia Signal Processing, pp. 320–323 (2006)
5. Cockshott, W.P., Balasuriya, S.L., Gunawan, I.P., Siebert, J.P.: Image enhancement using vector quantisation based interpolation. In: Proceedings of the British Machine Vision Conference (2007)
6. Coskun, B., Sankur, B.: Robust video hash extraction. In: Proceedings of the IEEE Signal Processing and Communications Applications Conference, pp. 292–295 (2004)
7. Daly, S.: Digital Images and Human Vision, chap. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, pp. 179–206. MIT Press, Cambridge (1993)
8. Dosselmann, R.: An evaluation of existing and emerging digital image and video quality metrics. Master's thesis, University of Regina, Regina, Saskatchewan, Canada (2006)
9. Dosselmann, R., Yang, X.D.: A prototype no-reference video quality system. In: Proceedings of the Canadian Conference on Computer and Robot Vision, pp. 411–417 (2007)
10. Dosselmann, R., Yang, X.D.: An empirical assessment of the structural similarity index. In: Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering, pp. 112–116 (2009)
11. Gonzalez, R., Woods, R.: Digital Image Processing. Prentice Hall, Upper Saddle River (2002)
12. Lay, D.C.: Linear Algebra and Its Applications. Addison Wesley Longman, Reading (1997)
13. Lewis, R.: Practical Digital Image Processing. Ellis Horwood, Chichester (1990)
14. Lubin, J.: Vision models for target detection and recognition, chap. 10. A Visual Discrimination Model for Imaging System Design and Evaluation, pp. 245–283. World Scientific, Singapore (1995)
15. Mai, Z.Y., Yang, C.L., Kuang, K.Z., Po, L.M.: A novel motion estimation method based on structural similarity for h.264 inter prediction. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp. 913–916 (2006)
16. Miyahara, M., Kotani, K., Algazi, V.R.: Objective picture quality scale (pqs) for image coding. IEEE Trans. Commun. **46**, 1215–1226 (1998)
17. Reibman, A.R., Poole, D.: Characterizing packet-loss impairments in compressed video. In: Proceedings of the IEEE International Conference on Image Processing vol. 5, pp. 77–80 (2007)
18. Rouse, D.M., Hemami, S.S.: Analyzing the role of visual structure in the recognition of natural image content with multi-scale ssim. In: Proceedngs of the SPIE Human Vision and Electronic Imaging Conference, vol. 6806 (2008)
19. Rouse, D.M., Hemami, S.S.: Understanding and simplifying the structural similarity metric. In: Proceedings of the IEEE International Conference on Image Processing, pp. 1188–1191 (2008)
20. Seshadrinathan, K., Bovik, A.C.: New vistas in image and video quality assessment. In: Proceedings of the SPIE Human Vision and Electronic Imaging Conference, vol. 6492 (2007)
21. Seshadrinathan, K., Bovik, A.C.: Unifying analysis of full reference image quality assessment. In: Proceedings of the IEEE International Conference on Image Processing, pp. 1200–1203 (2008)
22. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. **15**(11), 3440–3451 (2006)
23. Shnayderman, A., Gusev, A., Eskicioglu, A.M.: A multidimensional image quality measure using singular value decomposition. In: Proceedings of the SPIE Image Quality and System Performance Conference, vol. 5294, pp. 82–92 (2004)
24. Sonka, M., Hlavac, V., Boyle, R.: Image Processing, Analysis and Machine Vision. Brooks/Cole, Belmont (1999)
25. Sung, C.C., Ruan, S.J., Lin, B.Y., Shie, M.C.: Quality and power efficient architecture for the discrete cosine transform. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **E88A**(12), 3500–3507 (2005)
26. Süsstrunk, S., Winkler, S.: Color image quality on the internet. In: Proceedings of the SPIE Electronic Imaging Conference on Internet Imaging, vol. 5304, pp. 118–131 (2004)

27. Triola, M.F.: Elementary Statistics. Pearson, Boston (2005)
28. Vorren, S.S.: Subjective quality evaluation of the effect of packet loss in high-definition video. Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway (2006)
29. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process. Lett. **9**(3), 81–84 (2002)
30. Wang, Z., Bovik, A.C.: Why is image quality assessment so difficult? In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 4, pp. 3313–3316 (2002)
31. Wang, Z., Bovik, A.C.: Mean squared error: love it or leave it? IEEE Signal Process. Mag. **26**(1), 98–117 (2009)
32. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
33. Wang, Z., Bovik, A.C., Simoncelli, E.P.: Handbook of Image and Video Processing, 2 edn, chap. 8.3: Structural Approaches to Image Quality Assessment, pp. 961–974. Academic Press, New York (2005)
34. Wang, Z., Li, Q., Shang, X.: Perceptual image coding based on a maximum of minimal structural similarity criterion. In: Proceedings of the IEEE International Conference on Image Processing, vol. 2, pp. 121–124 (2007)
35. Wang, Z., Lu, L., Bovik, A.C.: Video quality assessment based on structural distortion measurement. Signal Process. Image Commun. **19**(2), 121–132 (2004)
36. Wang, Z., Sheikh, H.R., Bovik, A.C.: The Handbook of Video Databases: Design and Applications, chap. 41: Objective Video Quality Assessment, pp. 1041–1078. CRC Press, Boca Raton (2003)
37. Wang, Z., Simoncelli, E.P.: Translation insensitive image similarity in complex wavelet domain. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. II, pp. 573–576 (2005)
38. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multi-scale structural similarity for image quality assessment. In: Proceedings of the IEEE Asilomar Conference on Signals, Systems and Computers, vol. 2, pp. 1398–1402 (2003)
39. Winkler S.: Digital Video Quality: Vision Models and Metrics. John Wiley & Sons Ltd., West Sussex (2005)