

# Journal of Electronic Imaging

JElectronicImaging.org

## Improved structural similarity metric for the visible quality measurement of images

Daeho Lee  
Sungsoo Lim

# Improved structural similarity metric for the visible quality measurement of images

Daeho Lee<sup>a,\*</sup> and Sungsoo Lim<sup>b</sup>

<sup>a</sup>Kyung Hee University, Humanitas College, 1732, Deogyong-daero, Giheung-gu, Yongin 17104, Republic of Korea

<sup>b</sup>Kyung Hee University, Department of Electronics and Radio Engineering, 1732, Deogyong-daero, Giheung-gu, Yongin 17104, Republic of Korea

**Abstract.** The visible quality assessment of images is important to evaluate the performance of image processing methods such as image correction, compressing, and enhancement. The structural similarity is widely used to determine the visible quality; however, existing structural similarity metrics cannot correctly assess the perceived human visibility of images that have been slightly geometrically transformed or images that have undergone significant regional distortion. We propose an improved structural similarity metric that is more close to human visible evaluation. Compared with the existing metrics, the proposed method can more correctly evaluate the similarity between an original image and various distorted images. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JEI.25.6.063015](https://doi.org/10.1117/1.JEI.25.6.063015)]

Keywords: visual quality assessment; improved structural similarity metric; sharpness comparison.

Paper 16324 received Apr. 23, 2016; accepted for publication Nov. 11, 2016; published online Dec. 7, 2016.

## 1 Introduction

It is crucial to assess objectively image qualities for image processing applications because the assessments can compare with results of other methods to evaluate the performance. For measuring the performance of image correction, compressing and enhancing methods, such as denoising, JPEG compression, super-resolution, and frame rate upconversion,<sup>1-7</sup> and almost all objective evaluation metrics do not completely agree with the perceived subjective visibility of humans, while subjective evaluation is usually too inconvenient, time-consuming, and expensive.<sup>8</sup>

The simplest and most widely used metrics are mean squared error (MSE) and peak signal-to-noise ratio (PSNR); MSE is computed by averaging the squared differences of two signals, and PSNR is the ratio between the maximum value (Max) of a signal and the MSE as follows:

$$\text{MSE} = \frac{1}{M} \sum_{i=1}^M (x_i - y_i)^2, \quad (1)$$

$$\text{PSNR} = 10 \log_{10} \left( \frac{\text{Max}^2}{\text{MSE}} \right), \quad (2)$$

where  $x_i$  and  $y_i$  are the element of two signals, and  $M$  is the number of elements, e.g., the elements in image signals indicate pixels and the number of pixels should be equal to  $M$ . However, the MSE and PSNR are not very well matched to perceived visible quality.<sup>9-13</sup> A lot of image quality assessment methods based on error sensitivity have been proposed,<sup>14-19</sup> and they use the human visual system (HVS), contrast sensitivity function, discrete cosine transform, wavelet transform, and so forth. However, the similarity errors assessed by them may quite differ with the loss of

qualities, so some distortions may be clearly visible but these errors are not clearly observed in them.<sup>8</sup>

Recently, structural similarity (SSIM) has typically been used to determine visible quality.<sup>8,20</sup> This is a full reference image quality assessment method and it indicates how much an image is similar to the original image. It has three main components, which are structure, illuminance, and contrast. However, the components, especially structure component, are highly sensitive to translation, scaling, and rotation of an image. This means that although when images are translated and rotated as little as an unrecognizable amount, the SSIM is sensitively decreased.<sup>21</sup> Moreover, it may overestimate images that have undergone regional distortions such as JPEG compression.

In this paper, we aim at developing an improved structural similarity metric to outperform the typical SSIM, which can be used to overcome potential drawbacks. The proposed metric uses an improved structure comparison, and additionally uses a sharpness comparison.

## 2 SSIM and Its Drawbacks

Since humans usually use contrast, color, and frequency changes in their image quality measures,<sup>22</sup> the SSIM uses the luminance, contrast, and structure comparison shown in Fig. 1.<sup>8,22</sup> The SSIM of two images  $\mathbf{x}$  and  $\mathbf{y}$  is defined by the combination  $f()$  of three components as follows:<sup>8</sup>

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = f[l(\mathbf{x}, \mathbf{y}), c(\mathbf{x}, \mathbf{y}), s(\mathbf{x}, \mathbf{y})], \quad (3)$$

where  $l$ ,  $c$ , and  $s$  are the luminance, contrast, and structure comparison functions, respectively, defined by

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (4)$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (5)$$

\*Address all correspondence to: Daeho Lee, E-mail: [nize@khu.ac.kr](mailto:nize@khu.ac.kr)

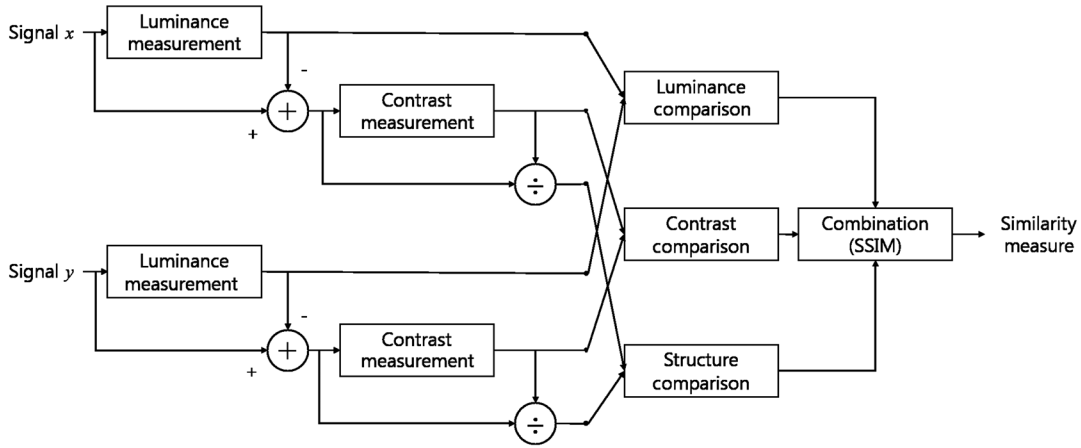


Fig. 1 Diagram of the SSIM measurement system.

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}, \quad (6)$$

where  $\mu_x$  and  $\sigma_x$  denote the mean and the standard deviation of  $\mathbf{x}$ ;  $\mu_y$  and  $\sigma_y$  denote the mean and the standard deviation of  $\mathbf{y}$ ;  $\sigma_{xy}$  denotes the covariance between  $\mathbf{x}$  and  $\mathbf{y}$ ; and  $C_1$ ,  $C_2$ , and  $C_3$  are constants used to avoid instability when the denominators are very close to zero. The values of  $l$ ,  $c$ , and  $s$  are in  $[0, 1]$  and they indicate higher similarities for each comparison function when the values are close to 1.

The local statistics are calculated within the local window having circular symmetric Gaussian weights, which are  $\mathbf{w} = \{w_i | i = 1, 2, \dots, N\}$  and  $\sum_{i=1}^N w_i = 1$  as follows:

$$\mu_x = \sum_{i=1}^N w_i x_i, \quad (7)$$



Fig. 2 Comparison of image similarity.

$$\sigma_x = \left[ \sum_{i=1}^N w_i (x_i - \mu_x)^2 \right]^{1/2}, \quad (8)$$

$$\sigma_{xy} = \sum_{i=1}^N w_i (x_i - \mu_x)(y_i - \mu_y), \quad (9)$$

where  $i$  is an index of the pixels in the Gaussian window and  $N$  is the total pixel number of the Gaussian window.

The combination of all comparisons between two images  $\mathbf{x}$  and  $\mathbf{y}$  is

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha \cdot [c(\mathbf{x}, \mathbf{y})]^\beta \cdot [s(\mathbf{x}, \mathbf{y})]^\gamma, \quad (10)$$

where  $\alpha > 0$ ,  $\beta > 0$ , and  $\gamma > 0$  are parameters used to adjust the relative importance. In order to simplify the expression and equalize the relative importance of the three components, they are generally set  $\alpha = \beta = \gamma = 1$  and  $C_3 = C_2/2$ , so we

also set the parameters in the same manner.<sup>8,21</sup> The results in a specific form of the SSIM index as follows:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (11)$$

To measure a single overall quality measure of the entire image, a mean SSIM (MSSIM) index is used as follows:

$$\text{MSSIM}(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{i=1}^M \text{SSIM}(\mathbf{x}_i, \mathbf{y}_i), \quad (12)$$

where  $\mathbf{X}$  and  $\mathbf{Y}$  are the original and the distorted images, respectively, and  $M$  is the number of pixels of images as used in Eq. (1).<sup>8</sup> MSSIM can be interpreted as a mean value of the SSIM index map.<sup>23</sup> Because SSIM values have the range of [0, 1], MSSIM also has the same range.

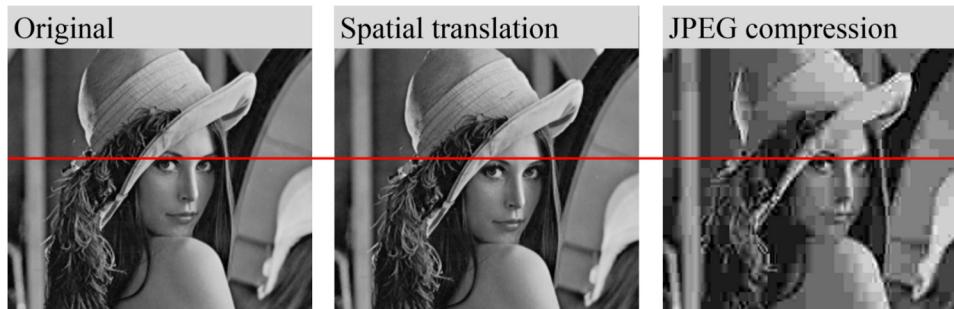


Fig. 3 Comparison of the original, ST, and JPEG compression image.

Table 1 Comparison of MSSIM and its components with MSSIM-S and its components about Fig. 3.

Images	MSSIM	$\bar{l}(\mathbf{x}, \mathbf{y})$	$\bar{c}(\mathbf{x}, \mathbf{y})$	$\bar{s}(\mathbf{x}, \mathbf{y})$	MSSIM-S	Mean of $s(\mathbf{x}, \mathbf{y})$	$\bar{h}(\mathbf{x}, \mathbf{y})$
ST	0.500	0.965	0.898	0.528	0.660	0.819	0.825
JPEG	0.706	0.995	0.917	0.771	0.640	0.822	0.822

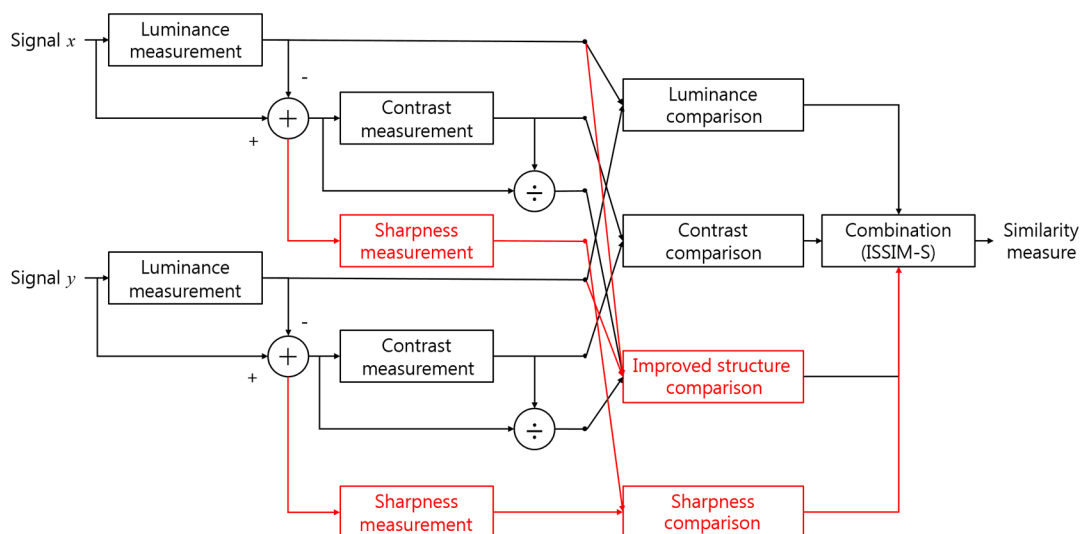


Fig. 4 Diagram of the proposed ISSIM-S measurement system.



The SSIM and MSSIM can be used to measure the similarity of two images. However, they have some drawbacks as shown in Fig. 2 and Table 2. First, images filtered by a low pass filter, such as a mean filter (MF), a median filter (MedF), and JPEG compression, are evaluated as having high similarity scores. Second, images that have been slightly distorted by some geometric transformations, such as spatial translation (ST) and rotation (RT), are evaluated as having low similarity scores.

### 3 New Structural Similarity

The main component of the SSIM that causes drawbacks is the structure comparison defined by Eq. (6). When we use Eq. (3) by only combining Eqs. (4) and (5), images that are

slightly geometrically transformed do not have low similarities as shown in Fig. 3 and Table 1, where  $\bar{l}(\mathbf{x}, \mathbf{y})$ ,  $\bar{c}(\mathbf{x}, \mathbf{y})$ , and  $\bar{s}(\mathbf{x}, \mathbf{y})$  are the mean of  $l(\mathbf{x}, \mathbf{y})$  in Eq. (4),  $c(\mathbf{x}, \mathbf{y})$  in Eq. (5), and  $s(\mathbf{x}, \mathbf{y})$  in Eq. (6). In Table 1,  $\bar{s}(\mathbf{x}, \mathbf{y})$  of the ST image is very low, while  $\bar{s}(\mathbf{x}, \mathbf{y})$  of the JPEG image is higher than that of the ST image. This example shows that the limitation of SSIM is sensitive to ST, scaling, and RT.

To reduce the weak effect of  $s(\mathbf{x}, \mathbf{y})$ , we define the structure comparison in a new way as follows:

$$\tilde{s}(\mathbf{x}, \mathbf{y}) = \frac{(2\sigma_{x-}\sigma_{y-} + C_2)(2\sigma_{x+}\sigma_{y+} + C_2)}{(\sigma_{x-}^2 + \sigma_{y-}^2 + C_2)(\sigma_{x+}^2 + \sigma_{y+}^2 + C_2)}, \quad (13)$$

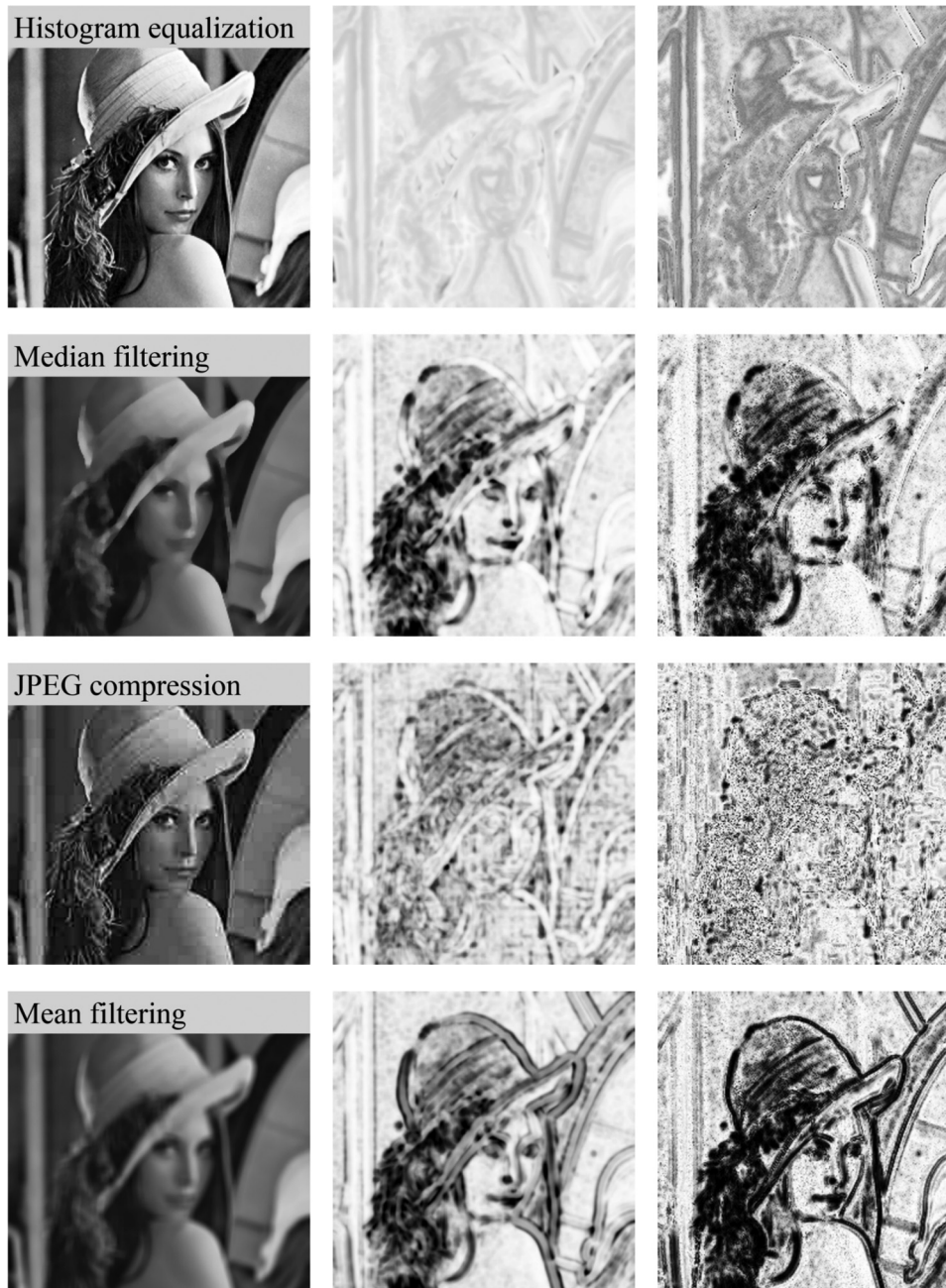
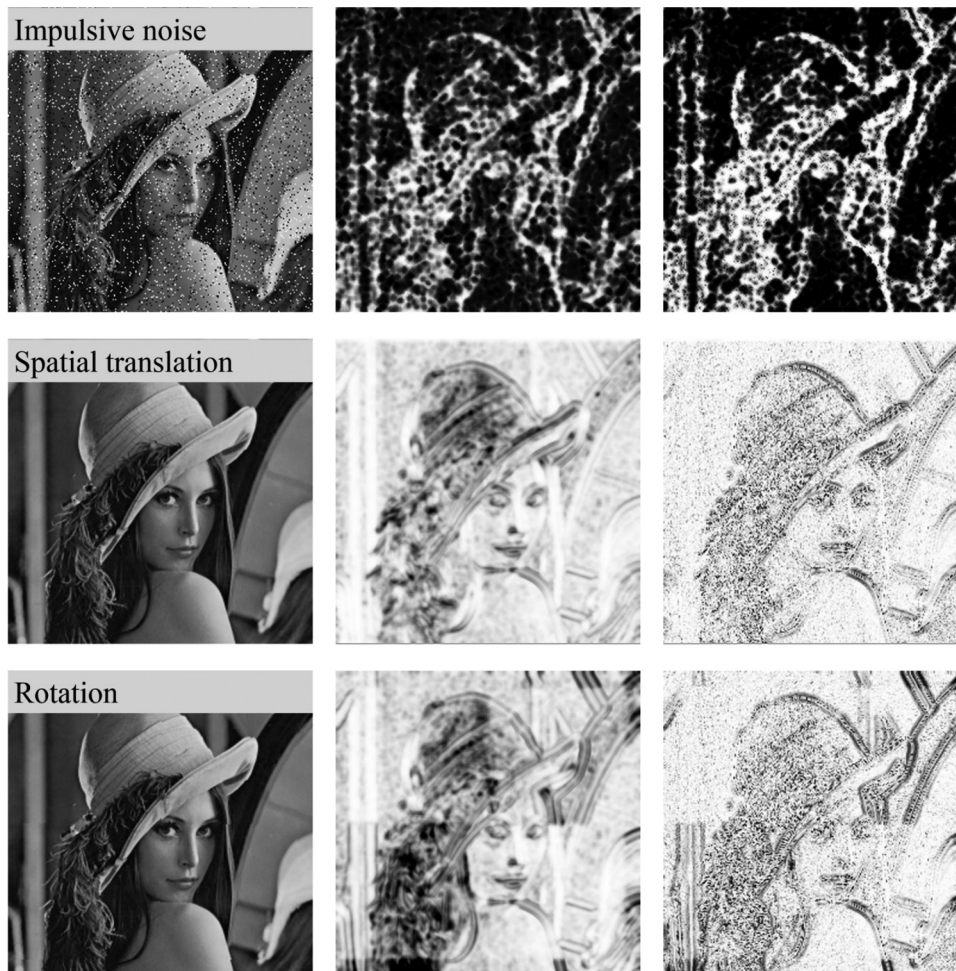


Fig. 5 Comparison of image similarity (from left to right: the evaluating images of Fig. 2, index maps of the SSIM, and index maps of the ISSIM-S).

where  $\sigma_{x-}$  and  $\sigma_{x+}$  denote the standard deviations for elements of  $\mathbf{x}$  smaller than and larger than  $\mu_x$ , respectively, and  $\sigma_{y-}$  and  $\sigma_{y+}$  denote the same for  $\mathbf{y}$ . In Ref. 8, structural information in an image is defined as those attributes that represent the structure of objects in the scene, independent of the average luminance and contrast, and structure comparison is conducted after luminance subtraction and variance normalization. So  $s(\mathbf{x}, \mathbf{y})$  is defined by the correlation between standard scores (z-score),<sup>24</sup>  $(x - \mu_x)/\sigma_x$  and  $(y - \mu_y)/\sigma_y$ . However, we define  $\tilde{s}(\mathbf{x}, \mathbf{y})$  as the correlation between standard deviations

for pixels having positive/negative standard scores because  $\sigma_{x-}$  and  $\sigma_{x+}$  can represent the structure of objects by dividing as locally brighter and darker regions. As shown in Fig. 3 and Table 1, the weak effect of  $s(\mathbf{x}, \mathbf{y})$  is relatively decreased compared to the original SSIM; however, the similarity of the ST image is lower than that of the JPEG image. That is to say, the SSIM still overestimates blurred images, when  $\tilde{s}$  is used as the structure comparison. Therefore, we add a new component, the sharpness comparison  $h(\mathbf{x}, \mathbf{y})$ , which is the correlation between the normalized digital Laplacian, defined as



**Fig. 6** Comparison of image similarity (from left to right: the evaluating images, index maps of the SSIM, and index maps of the ISSIM-S).



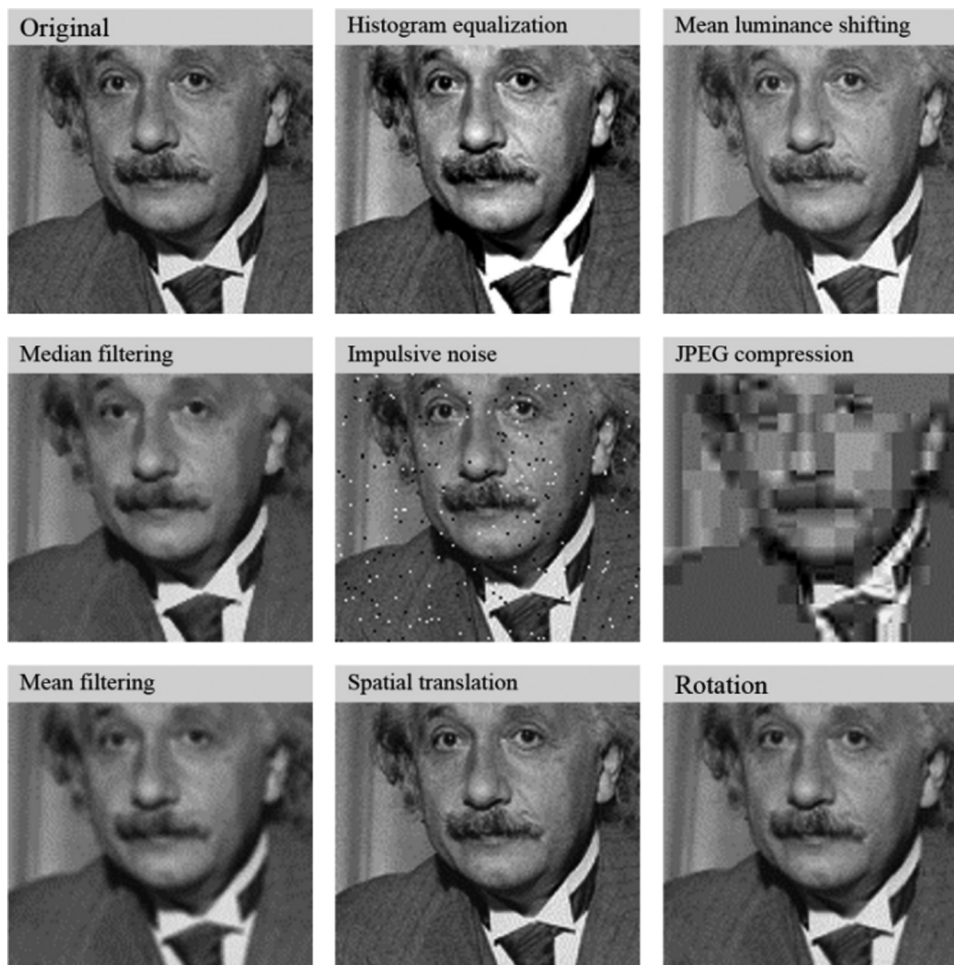
**Fig. 7** Comparison of image similarity (from left to right: the evaluating images, index maps of the SSIM, and index maps of the ISSIM-S).

**Table 2** Comparison of the PSNR, mean of the SSIM, mean of the ISSIM-S, and MOS rank of “Lena” image (the rank for each metric is shown in parentheses).

Images	PSNR	Mean of the SSIM	Mean of the ISSIM-S	MOS	MOS rank
HE	16.781 (6)	0.908 (1)	0.766 (5)	3.182	4
MLS	15.879 (8)	0.901 (2)	0.901 (1)	3.273	3
MedF	25.757 (3)	0.785 (5)	0.693 (6)	1.636	6
IN	16.098 (7)	0.297 (8)	0.313 (8)	1.545	7
JPEG	27.293 (1)	0.805 (4)	0.773 (4)	2.818	5
MF	23.888 (4)	0.711 (7)	0.623 (7)	1.273	8
ST	25.912 (2)	0.832 (3)	0.871 (2)	5.000	1
RT	23.474 (5)	0.759 (6)	0.832 (3)	4.909	2
$\rho$	0.048	0.595	0.881	—	—

**Table 3** Comparison of the PSNR, mean of the SSIM, mean of the ISSIM-S, and MOS rank of “Einstein” image (the rank for each metric is shown in parentheses).

Images	PSNR	Mean of the SSIM	Mean of the ISSIM-S	MOS	MOS rank
HE	23.278 (3)	0.924 (2)	0.795 (2)	3.583	4
MLS	23.028 (6)	0.986 (1)	0.986 (1)	4.417	2
MedF	29.536 (1)	0.827 (3)	0.746 (4)	2.333	7
IN	23.174 (5)	0.781 (5)	0.787 (3)	2.917	5
JPEG	23.237 (4)	0.557 (6)	0.447 (8)	1.000	8
MF	27.286 (2)	0.790 (4)	0.654 (7)	2.417	6
ST	18.729 (8)	0.393 (8)	0.680 (6)	4.833	1
RT	20.517 (7)	0.555 (7)	0.704 (5)	4.333	3
$\rho$	-0.643	-0.119	0.429	—	—



**Fig. 8** Comparison of “Einstein” image similarity (from left to right: the evaluating images, index maps of the SSIM, and index maps of the ISSIM-S).





**Fig. 9** Comparison of image similarity for different distortion levels (the numerics in parentheses indicate filter sizes of MF, quality factors of JPEG compression, and pixel amounts of ST).

**Table 4** Comparison of the PSNR, mean of the SSIM, and mean of the ISSIM-S for different distortion levels.

Images	PSNR	Mean of the SSIM	Mean of the ISSIM-S
MF (3 × 3)	29.280	0.896	0.822
MF (5 × 5)	25.725	0.792	0.697
MF (7 × 7)	23.888	0.711	0.623
JPEG (20)	29.936	0.871	0.844
JPEG (10)	27.701	0.806	0.772
JPEG (5)	25.109	0.706	0.640
ST (1)	25.912	0.832	0.871
ST (2)	21.881	0.690	0.806
ST (3)	20.060	0.607	0.754

$$h(\mathbf{x}, \mathbf{y}) = \frac{2|\nabla^2\mathbf{x}||\nabla^2\mathbf{y}| + C_2}{|\nabla^2\mathbf{x}|^2 + |\nabla^2\mathbf{y}|^2 + C_2}, \quad (14)$$

where  $\nabla^2\mathbf{x}$  and  $\nabla^2\mathbf{y}$  denote the normalized digital Laplacian given by

$$\nabla^2\mathbf{x} = \mathbf{x} - \mu_x. \quad (15)$$

The new similarity components  $s(\mathbf{x}, \mathbf{y})$  and  $h(\mathbf{x}, \mathbf{y})$  are satisfied with the properties for measurement metrics as follows:

1. Symmetry:  $S(\mathbf{x}, \mathbf{y}) = S(\mathbf{y}, \mathbf{x})$ ;
2. Boundedness:  $S(\mathbf{x}, \mathbf{y}) \leq 1$ ;
3. Unique maximum:  $S(\mathbf{x}, \mathbf{y}) = 1$ , if and only if  $\mathbf{x} = \mathbf{y}$ .

As shown in Fig. 4, the mean of  $h(\mathbf{x}, \mathbf{y})$  of the ST image is higher than that of the JPEG image. Finally, the improved SSIM which includes the sharpness comparison (ISSIM-S) can be defined as

$$\text{ISSIM-S} = l(\mathbf{x}, \mathbf{y}) \cdot c(\mathbf{x}, \mathbf{y}) \cdot \tilde{s}(\mathbf{x}, \mathbf{y}) \cdot h(\mathbf{x}, \mathbf{y}), \quad (16)$$

and the proposed ISSIM-S measurement system can be configured (Fig. 4).





Fig. 10 Comparison of image similarity for various scene contents.

To measure a single overall quality measure of the entire image, a mean ISSIM-s (MISSIM-S) index may be used as follows:

$$\text{MISSIM-S}(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{i=1}^M \text{ISSIM-S}(x_i, y_i). \quad (17)$$

The values of ISSIM-S and MISSIM-S are also in  $[0, 1]$  and these values indicate higher similarities when they are close to 1.

#### 4 Experimental Results

To evaluate the proposed similarity metric, which compares the PSNR and the SSIM, we tested some distorted images as shown in Fig. 2. In this test, we used an  $11 \times 11$  circular-symmetric Gaussian weight function, with a standard deviation of 1.5; normalized the unit sum equals to 1. The constants were selected to be  $C_1 = (0.01 \cdot 255)^2$ ,  $C_2 = (0.03 \cdot 255)^2$ , and  $C_3 = C_2/2$  as was done in Ref. 8.

These values seem somewhat arbitrary, but Wang et al. found that in their experiments, the performance of the SSIM index algorithm is fairly insensitive to variations of these values.

The local variance similarity between the original and the histogram-equalized images is quite different because histogram equalization (HE) is a nonlinear intensity transform. However, the SSIM is evaluated to have a high similarity score, while our new metric is evaluated as having a lower similarity than the SSIM. The ISSIM-Ss of the images, filtered by low pass filters, such as MF, MedF, and JPEG compression, are also evaluated to have lower similarities than the SSIM. In addition, the ISSIM-Ss of images that have been slightly geometrically transformed by ST and RT are higher than SSIMs. The results of the mean luminance shifting (MLS) and impulsive noise (IN) images show that the SSIMs and the ISSIM-Ss are evaluated with the same image but the result values are different.

To compare the different index maps of the SSIM and the ISSIM-S, the results of HE, MedF, JPEG, and MF are shown

**Table 5** Comparison of the PSNR, mean of the SSIM, and mean of the ISSIM-S for different scene contents.

Images	PSNR	Mean of the SSIM	Mean of the ISSIM-S
Goldhill (ST)	21.865	0.489	0.756
Goldhill (MF)	24.191	0.535	0.388
Goldhill (JPEG)	26.397	0.701	0.694
Boat (ST)	19.550	0.423	0.740
Boat (MF)	22.072	0.513	0.383
Boat (JPEG)	25.062	0.704	0.681

**Table 5 (Continued).**

Images	PSNR	Mean of the SSIM	Mean of the ISSIM-S
Airplane (ST)	20.141	0.664	0.801
Airplane (MF)	21.962	0.675	0.573
Airplane (JPEG)	26.356	0.805	0.751
House (ST)	24.755	0.676	0.839
House (MF)	26.362	0.766	0.636
House (JPEG)	30.557	0.825	0.759



**Fig. 11** Comparison of image similarity for various combinations of degradations.

in Fig. 5. The pixel values of the index map are normalized SSIM or ISSIM-S values. The index maps have different results, and the index maps of the ISSIM-S are darker than those of the SSIM because the MISSIM-Ss are lower than the MSSIMs. While the index maps of the ISSIM-S for IN, ST, and RT are brighter than those of the SSIM, because the similarities of the ISSIM-S are increased than those of the SSIM as shown in Fig. 6. The index maps of MLS are very similar as shown in Fig. 7.

To compare the mean opinion scores (MOSs), the rank of PSNR, mean of the SSIM, mean of the ISSIM-S, and MOS are shown in Table 2. To measure MOSs, we showed subjects the result images of each processing with the original image, and received their opinion scores, which have ranges of 1 (not similar) to 5 (very similar). Each comparison was implemented one-on-one with the original image and we randomized the order of the distorted images we showed to minimize order effects. The number of test subjects was 17 and none of them had any problems with their eyes. The experiments were implemented under the regulated illumination conditions and display conditions.

The scores themselves are subjective and not convincing but they can have meaning in relative comparison. Therefore, we used MOS ranks instead of MOS itself. The rank correlations by the MOS rank are also shown, where the rank correlation is computed by Spearman's rank correlation coefficient ( $\rho$ )<sup>25</sup> which is defined as follows:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (18)$$

where  $d_i$  denotes the difference of the  $i$ 'th rank and  $n$  denotes the ranking size. The rank correlation of the mean of the ISSIM-S is closer to 1 than the others.

We compared PSNR, SSIM, ISSIM-S, and MOS with another image shown in Fig. 8 and the results are shown in Table 3. The types of distortion are exactly the same as those of Table 2, but the only difference is the filter size. The resolution of test images in Table 2 is  $256 \times 256$  and the filter size is  $11 \times 11$ ; however, the resolution of test images in Fig. 8 is  $128 \times 128$  so we set the filter size as  $5 \times 5$ .

To evaluate the performance with different distortion levels, we tested a few more images: blurred images with different sizes of MF, images that have undergone various loss via JPEG compression, and images differently translated by ST (shown in Fig. 9 and Table 4). As the distortion level increases, PSNR, MSSIM, and mean ISSIM-S decrease, no matter the processing type. However, in ST, PSNR and MSSIM have the lowest values when it is translated only 3 pixels according to y axis, while mean ISSIM-S does not. ISSIM-S is also affected by translation but it is less sensitive than PSNR and SSIM methods.

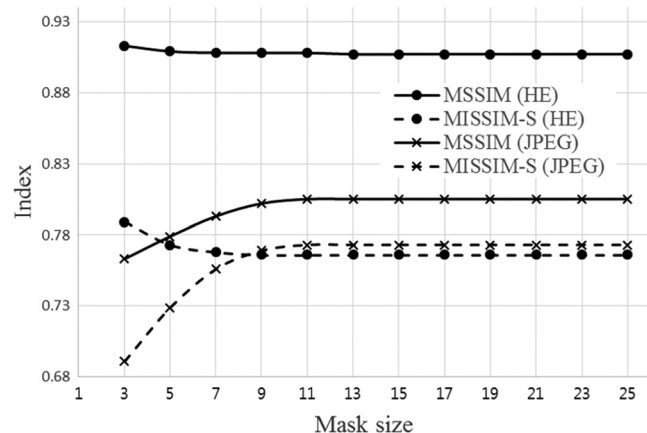
We conducted two additional experiments. First, comparison of ST, MF, and JPEG compression for various scene contents are shown Fig. 10 and Table 5. The resolutions of the tested images in this experiment are  $256 \times 256$ . The PSNR and the mean of SSIM values for each image are scored according to this order, ST < MF < JPEG. However, the mean of ISSIM-S shows another pattern, which is MF < JPEG < ST. The order of ISSIM-S is more reasonable than PSNR or SSIM. This result shows that the proposed image quality assessment method does not overestimate blurred images and

**Table 6** Comparison of the PSNR, mean of the SSIM, and mean of the ISSIM-S for various combinations of degradations.

Images	PSNR	Mean of the SSIM	Mean of the ISSIM-S
Goldhill (HE + IN)	11.581	0.409	0.235
Goldhill (ST + HE)	11.151	0.325	0.253
Goldhill (IN + ST)	17.659	0.390	0.490
Boat (HE + IN)	15.982	0.538	0.306
Boat (ST + HE)	14.300	0.250	0.329
Boat (IN + ST)	17.639	0.276	0.535
Airplane (HE + IN)	16.413	0.604	0.389
Airplane (ST + HE)	15.550	0.372	0.447
Airplane (IN + ST)	18.905	0.322	0.555
House (HE + IN)	16.378	0.394	0.185
House (ST + HE)	16.363	0.275	0.239
House (IN + ST)	20.125	0.365	0.439

it is much less sensitive to geometric transformations, which were one of the identified drawbacks of SSIM. Second, as shown in Fig. 11 and Table 6, we compared the PSNR, the mean of SSIM, and the mean of ISSIM-S for various combinations of degradations. The drawback of SSIM is that it is too sensitive to geometric translation and can be found when the degradations are combined. This result shows that MSSIM overvalues HE+IN while MISSIM-S evaluates moderately. It means that MISSIM-S is much closer to HVS because MISSIM-S is less sensitive to a small amount of geometric translation just as HVS is.

In addition, we tested the variations of MSSIM and MISSIM-S in terms of the size of the Gaussian window as shown in Fig. 12, where the  $11 \times 11$  window size is large



**Fig. 12** Variations of MSSIM and MISSIM-S in terms of the size of the Gaussian window.



enough because the variations are very small when the window size is larger than 11.

## 5 Conclusion

In this paper, we have proposed an improved structural similarity metric using structure and sharpness comparison functions to overcome the drawbacks of the SSIM metric. The structure comparison used segmented standard deviations by the mean, and sharpness comparison used the normalized digital Laplacian. The proposed metric can evaluate geometric transformed images with high similarities and cannot overestimate blurred images such as JPEG compression. The experimental results indicate that our similarity metric is superior to existing methods in respect to the perceived visibility of humans. Therefore, our method can be used to evaluate the performance of various methods such as image enhancement, frame rate upconversion, image compression, super-resolution, and image restoration.

## Acknowledgments

This research was partly supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01059091), and Institute for Information and communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. B0101-16-0033, Research and Development of 5G Mobile Communications Technologies using CCN-based Multi-dimensional Scalability).

## References

- U. S. Kim and M. H. Sunwoo, "New frame rate up-conversion algorithm with low computational complexity," *IEEE Trans. Circuits Syst. Video Technol.* **24**(3), 384–393 (2014).
- R. V. Babu, S. Suresh, and A. Pekis, "No-reference JPEG-image assessment using GAP\_RBF," *Signal Process.* **87**(6), 1493–1503 (2007).
- A. Shnaydeman, A. Gusev, and A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Trans. Image Process.* **15**(2), 422–429 (2006).
- W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.* **22**(2), 56–65 (2003).
- R. R. Schultz, L. Meng, and R. L. Stevenson, "Subpixel motion estimation for super-resolution image sequence enhancement," *J. Visual Commun. Image Represent.* **9**(1), 38–50 (1998).
- S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.* **9**(9), 1522–1531 (2000).
- A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, pp. 60–65 (2005).
- Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
- M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Process.* **70**(3), 177–200 (1998).
- A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.* **43**(12), 2959–2965 (1995).
- S. Winkler, "A perceptual distortion metric for digital color video," *Proc. SPIE* **3644**, 175–184 (1999).
- P. C. Eeo and D. J. Heeger, "Perceptual image distortion," *Proc. SPIE* **2179**, 127–141 (1994).
- Z. Qang, A.C. Bovik, and L. Lu, "Why is image quality assessment so difficult," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 3313–3316 (2002).
- W. Osberger, N. Bergmann, and A. Maeder, "An automatic image quality assessment technique incorporating high level perceptual factors," in *Proc. IEEE Int. Conf. Image Processing*, pp. 414–418 (1998).
- A. B. Watson, J. Hu, and J. F. McGowan III, "DVQ: a digital video quality metric based on human vision," *J. Electron. Imaging* **10**(1), 20–29 (2001).
- A. B. Watson et al., "Visibility of wavelet quantization noise," *IEEE Trans. Image Process.* **6**(8), 1164–1175 (1997).
- Y. K. Lai and C. C. J. Kuo, "A Haar wavelet approach to compressed image quality measurement," *J. Visual Commun. Image Represent.* **11**(1), 17–40 (2000).
- A. B. Watson, "DCT quantization matrices visually optimized for individual images," *Proc. SPIE* **1913**, 202–216 (1993).
- W. Xu and G. Hauske, "Picture quality evaluation based on error segmentation," *Proc. SPIE* **2308**, 1454–1465 (1994).
- Z. Wang and A.C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.* **9**(3), 81–84 (2002).
- Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 573–576 (2005).
- Y. A. Y. Al-Najjar and D. C. Soong, "Comparison of image quality assessment: PSNR, HVS, SSIM, UIQI," *Int. J. Sci. Eng. Res.* **3**(8), 1041–1045 (2012).
- Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process. Image Commun.* **19**(2), 121–132 (2004).
- K. Ma et al., "Objective quality assessment for color-to-gray image conversion," *IEEE Trans. Image Process.* **24**(12), 4673–4685 (2015).
- J. L. Myers and A. D. Well, *Research Design and Statistical Analysis*, 2nd ed., p. 508, Lawrence Erlbaum Associates, New Jersey (2003).

**Daeho Lee** received his MS and PhD degrees in electronics engineering from Kyung Hee University, Republic of Korea, in 2001 and 2005, respectively. He has been an associate professor in the Humanities College at Kyung Hee University, Republic of Korea, since 2005. His research interests include computer vision, pattern recognition, machine learning, image processing, image fusion, 3-D image reconstruction, computer games, ITS, HCI, electrical impedance tomography analysis, and digital signal processing.

**Sungsoo Lim** received his BS degrees in electronics and radio engineering and biomedical engineering and his MS degree in electronics and radio engineering from Kyung Hee University, Republic of Korea, in 2014 and 2016. He is currently pursuing his PhD in electronic engineering at the Kyung Hee University. His research interests include computer vision, image processing, intelligent transportation systems (ITS), human computer interaction (HCI), and medical image processing.