

# Assessment of random-noise contamination in digital images via testing on wavelet coefficients

SHARAD SILWAL, HAIYAN WANG\* AND DIEGO MALDONADO

Full-reference image quality assessment methods seek to measure visual similarity between two images (in practice, one original and the other its altered version). It has been established that traditional methods, such as Mean Square Error and Peak Signal-to-Noise Ratio poorly mimic the human visual system and much of the recent research in image quality assessment has been directed toward developing image similarity measures that are more consistent with assessments from human observers. Some extensively tested popular methods in this regard are Visual Image Fidelity (VIF), Structure Similarity Index (SSIM) and its variants Multi-scale Structure Similarity Index (MS-SSIM) and Information Content Weighted Multi-scale Structure Similarity Index (IW-SSIM). However, experiments show that these methods may produce drastically different similarity indices for different images contaminated with the same source of random noise. In this article, we propose a new full-reference image quality assessment method, namely, Wavelet-based Non-parametric Structure Similarity Index (WNPSSIM), specifically designed to detect visual similarity between images contaminated with all sorts of random noises. WNPSSIM is based on a rank test of the hypothesis of identical images conducted on the wavelet domain. Our experimental comparisons demonstrate that WNPSSIM provides similar ranking as MS-SSIM, IW-SSIM and VIF for images contaminated with different random noises in general though the methodology is very different. In addition, WNPSSIM corrects the aforementioned shortcoming of assigning sharply different similarity indices for different images contaminated with the same source of random noise.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 68U10, 97K80, 62H35; secondary 62G10.

KEYWORDS AND PHRASES: Image structure similarity, Non-parametric hypothesis testing, Full-reference, Human visual system (HVS), Discrete wavelet transform (DWT).

## 1. INTRODUCTION

Digital imaging has found massive applications in many branches of science: astronomy, meteorology, seismology, in-

dustrial inspection, aerial reconnaissance, autonomous navigation, to name but a few. As such, a reliable image quality index is highly desirable and vigorously pursued by the research community. In the literature, we find three generic approaches to an image quality assessment method; namely, full-reference, reduced-reference and no-reference. A full-reference image quality assessment method measures the visual similarity of a test image with respect to a supplied reference image. It is applicable when a true image or a standard reference image is known, such as in image compression and image coding (Richter and Larabi, 2008).

Many full-reference image quality assessment methods can be found in the literature. Multi-scale Structure Similarity Index (MS-SSIM) (Wang et al., 2003), Visual Image Fidelity (VIF) (Sheikh and Bovik, 2006), and Information Content Weighted Multi-scale Structure Similarity Index (IW-SSIM) (Wang and Li, 2011) are popular image quality assessment methods which are extensively tested and widely recognized to be consistent with the HVS. Others are Wang et al. (2004), Wang and Simoncelli (2005a), Sheikh et al. (2005a), Ninassi et al. (2006), Chandler and Hemami (2007), Gao et al. (2009) and Wang et al. (2011). We found that the extent to which these methods are capable of seeing through random noises in the same way the human visual system (HVS) does is limited. As a result, they may assign drastically different similarity index values for images contaminated with the same source of random noise. In practice, random noise is inevitably abundant in digital images. They occur at every stage of image acquisition and processing and they are quite difficult to remove completely. Thus, some level of random noise is always assumed in almost all forms of digital images. Moreover, if the contamination level of the random noise is low, the HVS easily sees through it and recognizes the underlying true image. This property of the HVS is the main focus of interest in this paper as it concentrates on developing an image similarity assessment method which possesses the ability to see through moderate amounts of random noise in images in the same manner as the HVS does and is robust for all types of random noises.

Our experiments found that the ability of the aforementioned methods to see through noises depends heavily on the content of the images. Figure 1 shows that IW-SSIM and MS-SSIM may assign values that spread out in a big range for different images contaminated with the same noise type or level. This is certainly a weakness indicating that

\*Corresponding author.

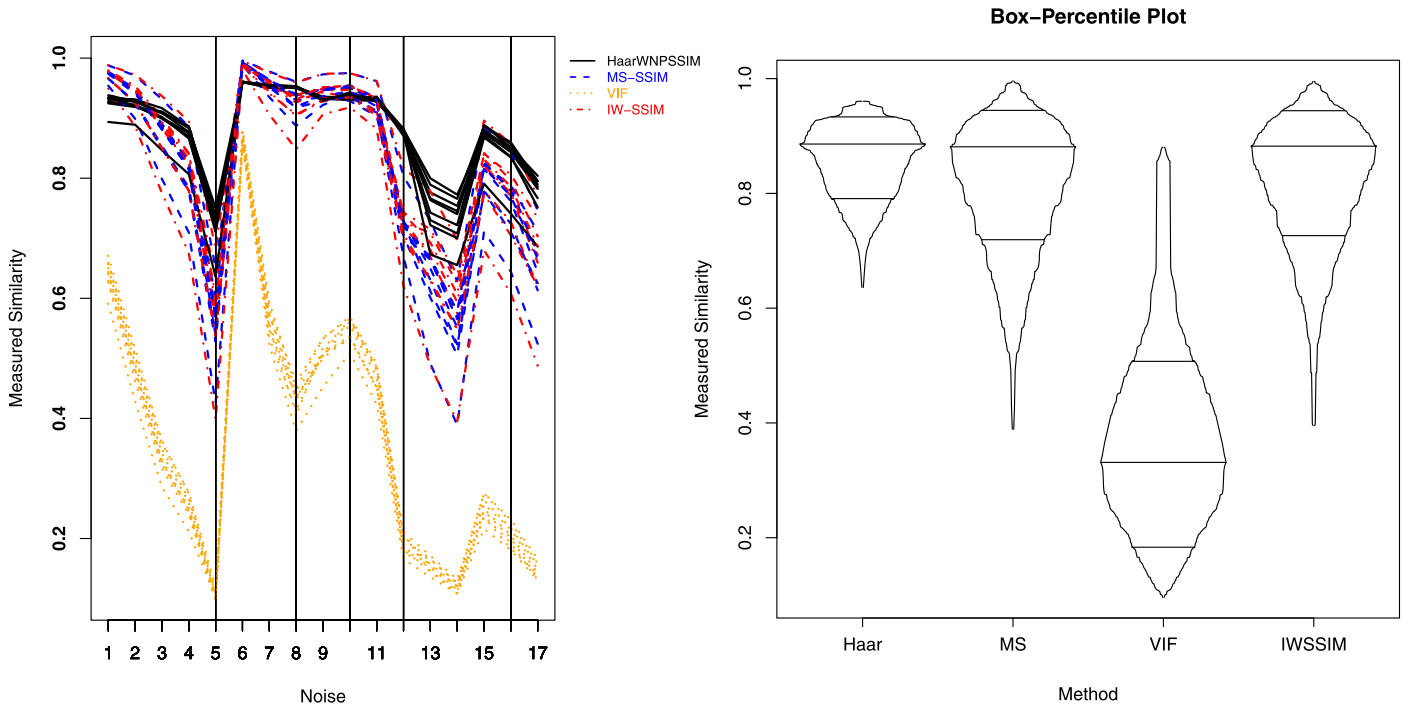


Figure 1. Visual illustration of the comparison of MS-SSIM, VIF, IW-SSIM and HaarWNPSSIM for 9 reference images with all of their respective 17 noisy versions which are described in Section 5.1 and can be viewed in Figures 8 and 9. Each curve in the left panel connects the similarity indices assigned by the same method for one image and its different noisy versions. Curves with the same color are for the same method on different reference images. The variation of the measured similarity for IW-SSIM and MS-SSIM for 5<sup>th</sup> noise contamination is much bigger than that for the 6<sup>th</sup> noise contamination.

these methods are somewhat lacking robustness to scene variation for noise assessment. In addition, the similarity assignment by these methods for the comparison between a reference image and its noisy version is seen to rely on the interaction between the image content and the noise. For instance, the performance of Structure Similarity Index (SSIM) (Wang et al., 2004), one of the hugely popular methods currently in use, is easily affected by the presence of random noises. It is seen to assign a higher similarity index value for the comparison of two completely different visually noise-free images than for that of the two images where one is visually noise-free and the other is its random noise contaminated version. In Figure 2 with  $A_4$  (lady with black hair) as the reference image, the test image  $A_8$  (peppers) is assigned a SSIM value of 0.2457 whereas the image  $N_{14,4}$  is assigned a SSIM value of 0.1885. This is clearly not reasonable since the latter, being only a random noisy version of the reference image, looks visually much closer. In addition, SSIM is unable to focus on the natural image content in the presence of random noise. For example, when  $N_{14,4}$  is the reference image, SSIM judges that  $N_{14,4}$  is much closer to  $N_{14,8}$  (noisy peppers) than to its own contamination-free version  $A_4$  since it assigned similarity index values of 0.73 and 0.1885 for these comparisons respectively. Visual Image Fidelity (VIF) introduced in Sheikh and Bovik (2006) has a similar drawback in addition to the fact that it often under-

estimates similarities between like images. The objective of the present work is to build an image similarity index which can assign consistent values on images with random noise contamination visible to the eye without being adversely affected by the contents in images.

We will first briefly list typical random noises that are common in digital images, since identifying the nature of these noise sources is critical to the development of quality assessment methods under realistic assumptions. The fact that different noises follow different distributions motivates us to propose an image quality assessment method based on non-parametric statistics. We will then review some image quality assessment methods with brief discussion of their relevance to the context of this paper and the reason they may yield quality indices inconsistent with the HVS. Then we will describe our new wavelet-based image similarity assessment method which is developed through a non-parametric test on hypotheses related to the wavelet coefficients of the reference and test images. We illustrate with experiments that the proposed image similarity index WNPSSIM provides a more realistic similarity assessment of images contaminated with random noise.

## 2. RANDOM NOISES IN IMAGES

Random noise is an inherent component of digital images. Natural images as we see in the open environment contain


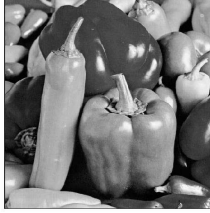


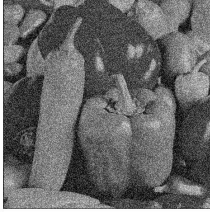

A <sub>4</sub>	A <sub>8</sub>	N <sub>2,1</sub>	N <sub>7,2</sub>	N <sub>14,8</sub>	N <sub>14,4</sub>
					
Reference Image	SSIM = 0.2457 MS-SSIM = 0.0052 VIF = 0.0314 IW-SSIM = 0.0302* WNPSSIM = 0.0005	SSIM = 0.1823 MS-SSIM = 0.0028 VIF = 0.0323 IW-SSIM = 0.0518* WNPSSIM = 0.0387	SSIM = 0.1083 MS-SSIM = 0.0004 VIF = 0.0336 IW-SSIM = 0.0543* WNPSSIM = 0.0338	SSIM = 0.0422 MS-SSIM = 0.0013 VIF = 0.0306 IW-SSIM = 0.0132* WNPSSIM = 0.2882	SSIM = 0.1885 MS-SSIM = 0.3889 VIF = 0.1081 IW-SSIM = 0.3952 WNPSSIM = 0.7726
SSIM = 0.1885 MS-SSIM = 0.3889 VIF = 0.0223 IW-SSIM = 0.3996 WNPSSIM = 0.7726	SSIM = 0.0414 MS-SSIM = 0.0024 VIF = 0.0047 IW-SSIM = 0.0172 WNPSSIM = 0.2578	SSIM = 0.0427 MS-SSIM = 0.0134 VIF = 0.0053 IW-SSIM = 0.0309* WNPSSIM = 0.1415	SSIM = 0.0343 MS-SSIM = -0.0004 VIF = 0.0057 IW-SSIM = 0.0298* WNPSSIM = 0.2917	SSIM = 0.7310 MS-SSIM = 0.0174 VIF = 0.4396 IW-SSIM = 0.1804 WNPSSIM = 0.1489	Reference Image

Figure 2. SSIM, MS-SSIM, VIF, IW-SSIM, and HaarWNPSSIM similarity indices for various test images with one clean image (A<sub>4</sub>) and one noisy image (N<sub>14,4</sub>) as reference images. Values marked with \* are absolute values of complex indices.

analog signals. Both the domain where the signals exist and the range where they take values are virtually continuous. Digital images are produced by devices with sensors that acquire these analog signals, process them and store them in computer-readable digital format. Thus, they contain signals on a discrete pixel domain and the pixels take values from a discrete set of possibilities. The two processes of discretization, that of the domain and of the range of pixels, are known as sampling and quantization, respectively. Random noise occurs naturally in this process of analog-to-digital conversion. Moreover, faulty devices and their faulty processing are bound to produce additional random noise in images. If the noise is independent of the signal it is referred to as white noise whereas pink and red noises refer to the noises that depend on the signal frequency. Sensors used in contemporary imaging devices such as complementary metal-oxide-semiconductor and charged-coupled device produce noise that are intensity-dependent, i.e., noise level varies with pixel intensity. We refer the reader to (Bovik, 2009) for more details on such noises as well as other common types of random noises in digital images, of which a brief summary is provided below.

- Gaussian noise. Also referred to as white noise or thermal noise or Johnson noise in other contexts, Gaussian noise is one of the most common noises found in digital images. See the images on the second column in Figure 8 for the effect of the same Gaussian noise level on different images and the first row of Figure 9 for an illustration of different levels of this noise on the same image. Some non-Gaussian noises similar to static in the AM radio can also be imitated by a Gaussian mixture model. The probability density function of a mixture model is of the form  $p_{\mathbf{q}}(x) = \alpha p_{\mathbf{q}_1}(x) + (1 - \alpha)p_{\mathbf{q}_2}(x)$  ( $0 < \alpha < 1$ ) for the noise that comes from two sources

having probability density functions  $p_{\mathbf{q}_1}(x)$  and  $p_{\mathbf{q}_2}(x)$  respectively, where the first occurs  $100\alpha\%$  of the time and the second occurs  $100(1 - \alpha)\%$  of the time.

- Poisson noise. There is a noise built into the detector of light particles or the photon counter of any image acquisition device. Since the random nature of photon-counting is governed by Poisson statistics, this noise is called Poisson or photon-counting or shot noise. A typical context of occurrence of Poisson noise is biomicroscopic images where live samples are often observed at very low light levels, due to acquisition-time and phototoxicity constraints (Luisier et al., 2010; Vonesch et al., 2006). See the images on the fourth column in Figure 8 and the image on the second row and fourth column of Figure 9.
- Salt and pepper noise. This is an impulse noise which is very common in digital images. Images corrupted with salt and pepper noise are degraded in only a few random pixel locations. This noise is often the result of faulty memory locations and signal transmission through noisy digital links. See the images on the third column in Figure 8 and the first three images on the second row of Figure 9.
- Speckle noise. Another noise that is more common in geoscience imaging such as active radar, synthetic aperture radar and biomedical imaging such as optical coherence tomography and ultrasound images is called speckle noise. This noise is produced in the imaging system where there are coherent light sources such as lasers or radar transmitters. The source of this noise is attributed to random interference of the coherent returns issued from the numerous scatterers present on a surface, on the scale of a wavelength of the incident radar wave (Gagnon and Jouan, 1997). In other words,

due to the microscopic variations of the surface roughness within one pixel, when the light is reflected off the surface, the received signal is subjected to random variations in phase and amplitude (Bovik, 2009). See the images on the fifth column in Figure 8 and the last two images on the second row and the first one on the third row of Figure 9.

### 3. SOME COMMON IMAGE SIMILARITY ASSESSMENT METHODS AND THEIR POSSIBLE LIMITATIONS

Images get degraded in time naturally or get corrupted through various processes they are subject to since they are captured. An image similarity assessment method is a measure of the degradation or corruption of an image. Full-reference image similarity assessment methods measure the loss of quality in a noisy image relative to a supplied reference image which is treated as the pure or ideal version. In this section, we define and discuss some common full-reference image similarity assessment methods. Let  $X$  be an original or reference image and  $Y$  be its estimate or a test image, both 8-bit images of size  $m \times n$ .

#### 3.1 Traditional similarity indices

Most conventional image similarity indices are based on the Euclidean distance or the  $L^2$ -norm because it is regarded as the natural distance in most spaces. Let  $X$  and  $Y$  be images of size  $m \times n$ . The mean squared error (MSE) of  $Y$  with respect to  $X$ , the error in terms of the Euclidean distance, is defined as

$$\text{MSE}(X, Y) = \frac{1}{mn} \|X - Y\|_2^2 = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n |X(i, j) - Y(i, j)|^2.$$

The human eye is regarded to be the true benchmark for image similarity assessment methods which is known to process an image in a much more complicated sense, viewing image pixels en bloc to capture visual similarity. As is clear from the above definition, MSE merely takes into account each individual pixel value without any adjusted weighting in measuring the similarity between images. As a result, MSE turns out to be a very poor indicator of visual similarity between images. This has been discussed in great detail along with extensive illustrations in Wang and Bovik (2009). A popular traditional measure of image similarity is Signal-to-Noise Ratio (SNR) defined as

$$\text{SNR}(X, Y) = 10 \log_{10} \left( \frac{\|X\|_2^2}{\|X - Y\|_2^2} \right).$$

It is measured in decibel (dB) units. A larger value of SNR in general indicates the presence of less noise in the image. Since SNR is defined through MSE, it shares the drawbacks of MSE as an image similarity measure. Since MSE, Total

Variation (TV), and other similar measures such as  $L^1$ -norm which are purely mathematical in nature treat each pixel identically, we avoid reiteration of them.

#### 3.2 Structure similarity index (SSIM)

Introduced first in Wang and Bovik (2002) and fully developed in Wang et al. (2004), a correlation-based image similarity measure called the Structure Similarity Index (SSIM) emerged as a solid method giving more realistic image similarity measures than  $L^2$ -norm-based methods such as MSE and Signal-to-Noise Ratio (SNR). The introduction of SSIM led to many other full-reference image similarity measures in the literature seeking to eliminate some of the limitations of SSIM. Notwithstanding its drawbacks, SSIM is found widely cited in the literature owing to its thorough construction which incorporates several factors of visual perception.

SSIM( $X, Y$ ) is defined to be the average of all local indices SSIM( $x, y$ ),  $x = X(i, j)$ ,  $y = Y(i, j)$ ,  $i = 1, \dots, m - 10$ ,  $j = 1, \dots, n - 10$ , which, in turn, is the product of three factors:

$$\begin{aligned} \text{SSIM}(x, y) &= l(x, y) \cdot c(x, y) \cdot s(x, y) \\ &= \left( \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right) \cdot \left( \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right) \cdot \left( \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \right), \end{aligned}$$

where  $l(x, y)$ ,  $c(x, y)$  and  $s(x, y)$  are the local luminance comparison, local contrast comparison and local structure comparison or correlation between the images  $X$  and  $Y$  respectively. For local statistics on each pixel, a  $11 \times 11$  sub-matrix of  $X$  with  $X(i, j)$  as the (1,1) entry is chosen and the local mean  $\mu_x$  is computed for this local window as a weighted mean with the same size Gaussian low-pass filter with standard deviation 1.5 as the weight matrix. Similarly, the local mean  $\mu_y$ , local standard deviations  $\sigma_x$  and  $\sigma_y$  and the local covariance  $\sigma_{xy}$  are computed.  $C_1 = 0.01$ ,  $C_2 = 2C_3 = 0.03$  are small positive constants incorporated to avoid numerical instability. Note the cancelation in the formula for SSIM( $x, y$ ) after the choice  $C_2 = 2C_3$ . The SSIM( $X, Y$ ) described here is the simplest and default version of the SSIM MATLAB code as made available by Wang et al. (2004). We refer the reader to the original paper for details on a more complete version of SSIM.

The local statistics in SSIM was necessary to account for local variations of image features across a scene and also due to the fact that at typical viewing distances, only a local area in the image can be perceived with high resolution by the human observer at one time instance. The local smoothing with a low-pass filter was reported to eliminate undesirable ‘‘blocking’’ artifacts and exhibit a locally isotropic property in the SSIM index map (Wang et al., 2004). Such local smoothing makes SSIM quite insensitive to boundary or edge loss. This, however, becomes unfavorable when the loss is too much to preserve the visual quality. The third factor in the SSIM formula is Pearson’s correlation which is too naive a measure for structure similarity comparison



since it only quantifies the strength of a linear relationship. It is well-known that Pearson’s correlation is a valid measure of dependence between two variables only if one variable is a linear function of the other, say  $Y = X\beta + \epsilon$ , and the noise  $\epsilon$  has a Gaussian distribution. When the noise distribution is not Gaussian, it is possible that the correlation between  $X$  and  $Y$  is zero but  $X$  and  $Y$  are perfectly related (see Wang et al. (2010) for explicit examples). Indeed, we have found in our experiments that SSIM yields persistently high scores for blurred distortions and deceptively low scores for slight random noisy versions in relation to how the human eye perceives the quality loss.

### 3.3 MS-SSIM, VIF and IW-SSIM

Among all the full-reference image quality assessment methods in the literature, the three multi-scale decomposition methods MS-SSIM (Wang et al., 2003), VIF (Sheikh and Bovik, 2006) and IW-SSIM (Wang and Li, 2011) are most relevant to our article. In most practical purposes (not always, though, see Figure 2), they have similarity indices on a scale of 0 to 1 (perfect similarity) just like the proposed image quality assessment method in this article. We refer the reader to the original papers for more detail and only provide here a very brief description of these methods.

MS-SSIM was introduced by some of the co-developers of SSIM as an improvement over SSIM. The methodology involves a multi-scale decomposition of involved images through an iterative process of applying a low-pass filtering and downsampling by 2 in each dimension. It is apt to mention that the standard discrete wavelet transform is very similar to this process as it also applies a low-pass filtering associated with a particular wavelet followed by downsampling by 2. If the low-pass filters associated with a particular wavelet are of length  $2N$ , a one-dimensional signal of length  $n$  will reduce to an approximated signal of length equal to the integer less than or equal to  $(n - 1 + 2N)/2$ . The final MS-SSIM index is constructed by combining the three components of SSIM, namely, luminance, contrast and structure computed at different scales. As is the case with SSIM, due to a correlation factor as its component, the values of the MS-SSIM index, in principle, might range between  $-1$  and  $1$ .

VIF, proposed by two of the co-developers of SSIM, uses a statistical approach developed by the authors in Sheikh et al. (2005a) where the so-called natural images are modeled in the wavelet domain as perceived by the HVS. A more detailed discussion of the natural images in the wavelet domain is provided in Section 4.1. First, the source, distortion and HVS models based on the Gaussian Scale Mixture (GSM) model (Wainwright et al., 2001) are developed. Reference images are then modeled as outputs of a natural source having passed through the HVS channel. Distorted images pass through an additional distortion channel before entering the HVS channel. Next, the image information is computed as the mutual information between the input and

the output of the HVS channel. Finally, the VIF index is computed as the ratio of the distorted image information and the reference image information. VIF is not a symmetric index and can take on values bigger than 1 for contrast-enhanced test images. However, for most practical purposes, including the one considered in this paper which restricts the attention to only random-noisy test images, VIF does lie between 0 and 1.

IW-SSIM is another recent image quality assessment algorithm proposed as an improvement over MS-SSIM. Most methods, including MS-SSIM, provide elaborate ways of measuring local image similarity and a relatively more simplistic method of pooling those local measurements into a global image similarity measure. The main feature of IW-SSIM is a more sophisticated way of pooling together local measurements. The authors provide an elaborate formula to compute weights proportional to local information content based on the image information measure in the VIF index described above. The IW-SSIM index employs the weighted pooling of the SSIM components computed at different scales and locations. In the event of some unconventional comparison, the IW-SSIM index can also take complex values (see examples in Figure 2).

### 3.4 Other image similarity assessment methods

Development of full-reference image quality assessment methods in agreement with the HVS is one of the highly pursued topics of research in the image processing community. A discussion of several old and new full-reference image quality assessment methods can be found in Ponomarenko et al. (2009a) along with a comparison among them. One of the earlier methods based on the idea of a wavelet analysis was proposed in Lai and Kuo (2000). Besides those already mentioned, some other methods include Complex Wavelet-domain SSIM (CWSSIM) (Wang and Simoncelli, 2005b), Information Fidelity Criterion (IFC) (Sheikh et al., 2005a), Peak Signal-to-Noise Ratio human visual system (Egiazarian et al., 2006), Visual Signal-to-Noise Ratio (Chandler and Hemami, 2007) and Peak Signal-to-Noise Ratio taking into account Contrast Sensitivity Function, Between-coefficient Contrast Masking of Discrete Cosine Transform basis functions (Ponomarenko et al., 2009a).

In Wang et al. (2011), the authors of this article have proposed an image similarity assessment method called P-value-based Structure Similarity Index (PSSIM) which is constructed by applying a rank-based non-parametric test of independence developed in Wang et al. (2010). It has been shown in Wang et al. (2011) that PSSIM consistently outperforms SSIM and MSE as an image similarity index in the context of both random and deterministic noises. The emphasis of PSSIM is on structural loss of image features. Hence, it employs a test of independence which is violated when there is a loss of image structure. By contrast, the goal of this article is to develop an image similarity assessment

method with the emphasis on the ability to see through random noise at reasonable levels such that the human eye can still identify the image content of the underlying image. Hence, in the construction of the WNPSSIM index proposed in this article, we employ a completely different test that appropriately detects the structural difference in the presence of random noise in images by assessing whether or not the noise come from an identical source. Although both PSSIM and the proposed WNPSSIM method employ statistical tests, there is a significant methodological difference between the two methods. PSSIM employs a test in the pixel domain and WNPSSIM employs a test in the wavelet domain, where the HVS properties are known to be better captured. Owing to the pixel-domain methodology of PSSIM and the very specific objective of this paper, PSSIM is deemed irrelevant to the current context. Thus, we do not provide any more discussion of PSSIM in this paper and the reader is referred to the original article Wang et al. (2011) for its detailed description. Instead, we include in this article comparisons of the proposed method with MS-SSIM, VIF, and IW-SSIM (see Section 3.3) which are more relevant to our current context.

## 4. OUR PROPOSED IMAGE SIMILARITY ASSESSMENT METHOD

This section describes the details of our proposed image similarity assessment method. This method employs, in the wavelet domain, a non-parametric test developed by Wang et al. (2008). In Section 4.1, we briefly revisit the wavelet paradigm applied to natural images. A description of our construction of the proposed similarity index is provided in Sections 4.2 and 4.3.

### 4.1 The wavelet domain and the natural images in the wavelet domain

The wavelet transform provides a representation of images that enables a rich image analysis due to various properties that are not available in the raw image data, i.e., the data in the pixel domain. Moreover, the decay and localization properties of wavelets have made the wavelet domain even better than the Fourier domain where the representative profiles are global sinusoidal signals. We refer the reader to (Daubechies, 1992) and (Mallat, 2008) for a detailed treatment of the wavelet and multi-resolution approximation theories. An illustration of the wavelet transform of data in the context of images is given in Figure 3, where the image on the left is the data in the pixel domain and the image on the right is the visualization of the data in the wavelet domain which consists of the four sets of wavelet coefficients called subbands. In Figure 3, the approximation subband (ap), horizontal subband (h), vertical subband (v) and diagonal subband (d) are represented by the four blocks: the top left, top right, bottom left and bottom right respectively.

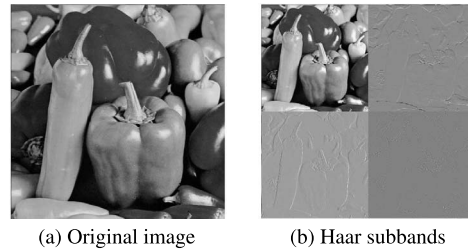


Figure 3. Original image and its Haar subbands at level 1.

Mathematically, all  $M \times N$  matrices with entries lying in a range of values representing grayscale intensities, most commonly in  $\{0, 1, \dots, 255\}$ , are (monochromatic) images. However, the majority of such images neither resemble anything in nature nor are encountered in most applications. The images that are mainly of interest to the image analysis community are the so-called *natural images*, that is, the images that occur in nature and everyday life. The collection of these natural images is believed to be a very thin subset of  $\mathbb{R}^{MN}$ . Indeed, these images typically have spatial structure mostly consisting of smooth areas interspersed with occasional edges (Simoncelli and Adelson, 1996). This structure, which distinguishes an image as a natural image, is better explained and exploited via some transformation in domains other than the pixel domain. In the Fourier domain, the power spectrum  $S(f)$  of natural images typically possess the decay property  $S(f) = Cf^{-\gamma}$  where  $f$  is the spatial frequency,  $C$  is some constant and  $\gamma \approx 2$  (Field, 1987; Field and Brady, 1997; Millane et al., 2003; Ruderman, 1997; Tolhurst et al., 1992). Likewise, natural images when transformed into the wavelet domain are also seen to enjoy a number of properties. One is the sparsity of the wavelet coefficients, i.e., most wavelet coefficients are near-zero and can be discarded for approximation purposes (Mallat, 2008). Second, the sizes of the detail wavelet coefficients can be modeled by the following generic probability density function:  $h(u) = Ke^{-(|u|/\alpha)^\beta}$  where  $\alpha$  and  $\beta$  are the decay and variance parameters respectively and  $K$  is a normalizing constant (Mallat, 1989).

As regards the goal of this paper which is to develop an image similarity method close to the human visual system (HVS) in performance, the most important property of the wavelet transform of the natural images is its high similarity with the HVS (Field, 1999; Ninassi et al., 2008; Xiong et al., 1997). The process of the HVS can be tersely explained in two phases: the pre-processing of the light reflected off an object by the neurons before it enters the visual cortex as an input and the post-processing thereof by the cortical cells eventually leading to object recognition by the human brain. During both phases, the different visual information components are processed by a series of different and independent channels which are sensitive to the visual stimuli with specific spectrum location, frequency and orientation (Gao et al., 2010). The wavelet decomposition of

an image exhibits striking similarity with the HVS regarding the independent processing of the different components of an image relative to location and orientation. This property of the HVS is referred to as *multi-channel parallel pathways*. Hence, the representation of images in the wavelet domain better captures this property of the HVS by sharing basic properties of neural responses in the primary visual cortex of mammals which are presumably adapted to efficiently represent the visually relevant features of images (Portilla et al., 2003). Thus, keeping in view that the discrete wavelet transformation establishes a multiresolution representation which attempts to mimic the decomposition performed by the cortical basis, all with relatively low complexity (Chandler and Hemami, 2005; Watson, 1987), we take the wavelet domain as our natural domain for image quality assessment.

## 4.2 The non-parametric test in the context of image similarity

Wang et al. (2008) have developed a non-parametric test which evaluates whether the observations in a vector or a signal are independently identically distributed, i.e., iid. This is a rank-based hypothesis test which is essentially an ‘‘iid noise detector’’. We refer the reader to Wang et al. (2008) and Wang and Akritas (2004) for details. Our work in this current paper is to measure the similarity between the two images, say, the reference image  $X$  and the test image  $Y$  by applying the above-mentioned test to the relative difference between them, namely, the ratio  $\frac{X-Y}{|X|+|Y|}$ . If  $Y$  is some random-noise contamination of  $X$ , this ratio measures how big the noise at each pixel location is relative to the images. The denominator may be replaced by  $X$  if an asymmetric similarity index is desired. By testing whether the local ratios are ‘‘iid’’, we obtain a similarity assessment method of how far  $Y$  departs from  $X$  locally and whether this departure is negligible.

We have discussed in Section 4.1 that the signal features relevant to the human visual system (HVS) are better captured by their transforms in the wavelet domain than in the raw pixel domain. Hence, instead of applying the test directly on the difference of the reference image  $X$  and the test image  $Y$ , we consider their four first-level subbands, namely, approximation subband (ap), vertical detail subband (v), horizontal detail subband (h) and diagonal detail subband (d) and apply the test to their relative differences, namely,

$$(1) \quad \frac{X_{ap} - Y_{ap}}{|X_{ap}| + |Y_{ap}|}, \quad \frac{X_v - Y_v}{|X_v| + |Y_v|}, \quad \frac{X_h - Y_h}{|X_h| + |Y_h|} \quad \text{and} \quad \frac{X_d - Y_d}{|X_d| + |Y_d|}.$$

A wavelet decomposition of the images in our experiment data was done in MATLAB using the DIPUM package (Gonzalez et al., 2009). A vast number of choices for wavelets is also available with the wavelet toolbox in MATLAB. In our experiment, we have used the five wavelets:

Haar, Daubechies 4, Symmlet 4, Jpeg 9.7 and Biorthogonal 6.8. All the five versions of WNPSSIM give a better performance in comparison to other methods. Since the similarity measures assigned by them are very close to one another, we take HaarWNPSSIM as the representative of all five versions and only refer to this one in most of our discussions. Further remarks on the types of wavelets and their performances are provided in Section 6.1.

Next, we adopt the same principle of incorporating local measurements as was done by SSIM. Instead of applying the test only once to each relative difference of subbands in (1), we consider multiple local windows in it and apply the test to each local window. For example, in a local window of  $\frac{X_{ap}-Y_{ap}}{|X_{ap}|+|Y_{ap}|}$ , a non-rejection of the test indicated by a large ‘p-value’ provides evidence to believe that the relative difference between the two subbands  $X_{ap}$  and  $Y_{ap}$  restricted to that local region is close enough to ‘iid’ random noise (subject to type II error), which, in turn, can be interpreted as the two subbands having a high level of structural similarity between them in that local window, except for some random noise. We then record the percentage of non-rejections of the hypothesis tests applied to the local windows within each subband image and we denote them by  $p_{ap}$ ,  $p_h$ ,  $p_v$  and  $p_d$  after approximation, horizontal detail, vertical detail and diagonal detail respectively.

We emphasize that  $p_{ap}$ ,  $p_h$ ,  $p_v$  and  $p_d$  are proportions of non-rejections, and not p-values. They have a more direct meaning in the sense that they represent the proportions of local regions of the two images that resemble each other structure-wise in corresponding subbands. This idea of considering the proportion of non-rejections was successfully implemented by the same authors in Wang et al. (2011) where another image similarity index based on a different rank-based non-parametric test in the spatial domain was proposed and shown to imitate the HVS better than other contemporary similarity indices, most notably, SSIM. This paper, as we have described above, seeks to focus on the seeing-through ability of random noises of the HVS. Also, note that the objective of our test is the non-rejections of the null hypotheses rather than the rejections which is different from the usual objective of a hypothesis test in most cases. Therefore, we do not recommend to use multiple-comparison adjustments aimed at controlling the family-wise error rate. Instead, a cut-off level of  $\alpha$  for non-rejection is more stringent than the threshold by Bonferroni corrected level ( $\approx \alpha$  divided by the number of tests). This ensures the p-values need to be large enough to qualify as a non-rejection. We have implemented in our experiment  $\alpha = 0.01$ .

All the images in our experiment described in Section 5.1 and partially illustrated in Figures 8 and 9 are of size  $512 \times 512$ . Each subband is one-fourth the original image size, i.e., of size  $256 \times 256$ . For these subbands, we consider local windows of size  $11 \times 11$ , much in the same spirit as that of SSIM. The application of the test on local windows provide local similarity measurements which are then

pooled together to obtain a global similarity measurement. In choosing the size for a shift of a local window so as to cover the whole image, there is a trade-off between the amount of local comparisons to take into account and the computational cost. A local window of size  $11 \times 11$  and its horizontal or vertical shift by only 1 pixel differ, at the most, at 11 observations out of a total of 121 observations. As a result, the p-values produced with the test applied with respect to these two local windows are almost identical. On the other hand, the intersection of the windows is also necessary, to some extent, to ensure that the spatial connection within the local regions of the two images is carried over to the final similarity measurement between them. Hence, in order to strike a balance, our rule is to shift the local windows horizontally and then vertically by half of the window width, namely, by 5 pixel locations.

With the proportions of non-rejections  $p_{ap}$ ,  $p_h$ ,  $p_v$  and  $p_d$  corresponding to the four subbands defined in this section, a similarity index between the reference image and the test image is calculated as a weighted geometric mean of these proportions as below:

$$\text{WSSIM} = p_{ap}^{0.95} \cdot p_h^{0.02} \cdot p_v^{0.02} \cdot p_d^{0.01}.$$

The weighted mean WSSIM is so constructed as to give proper credential to each subband to the extent it supplies structural information to the HVS. Note that all the proportions as well as weights are positive numbers less than one. Hence, raising the proportions of non-rejections from each subband to such weights has the effect of boosting those proportions. The construction of the weighted geometric mean WSSIM is based on the idea that a proportion that is to be given less importance is assigned a smaller weight and is boosted thereof. The weights, as assigned, reflect that WSSIM honors more of the conclusion from the approximation subband and pays less attention to the detail subbands. We list below a few points to justify our approach:

1. The approximation subband is a subsampled version of the original image and, thus, clearly retains most visual structure of the original image. On the other hand, the horizontal, vertical and diagonal subbands contain details of the original image missed by the approximation subband in the respective orientations. For reference, see Figure 3 which illustrates an original image and its four subbands of Haar wavelet coefficients. Consequently,  $p_{ap}$  needs to be boosted the least as the visual structure of the original image is clearly obvious in the approximation subband and  $p_h$ ,  $p_v$  and  $p_d$  need to be boosted more as the fine details in the detail subbands are not so visually obvious.
2. The percentage of structural similarity information captured by each subband will be different for different data of natural images. However, we believe that data-driven estimates of these percentages might not provide convincing evidence to prove that the weights as suggested

by such estimates will be any better. Instead, we rely on the fact that the empirical distribution of the size of the detail coefficients for natural images exhibit the pattern of double exponential distribution (Mallat, 1989). Due to this fact, the structural information captured by the detail coefficients is significantly understated. Hence, we need to boost the proportion of non-rejections from the detail subbands to a greater extent.

3. The use of weights can be viewed as an alternate way of thresholding wavelet coefficients which is a standard practice in image reconstruction. The idea of thresholding is based on the sparsity property of wavelet coefficients which says that it is sufficient to consider only a few wavelet coefficients and discard the rest of them which are near-zero. Wavelet thresholding has a very strong theoretical support provided in Donoho (1995). Most information about the underlying true image structure will be retained by thresholding because every wavelet coefficient contributes random noise, but only a very few wavelet coefficients contribute signal (Donoho and Johnstone, 1994). Since determining the optimal cut-off for thresholding is computationally extensive (Kim and Akritas, 2010), we instead keep 25% of the total wavelet coefficients by assigning the weight 0.95 to the approximation subband (keeping 95% of the approximation coefficients which account for 1/4 of the total) and utilize the rest of the coefficients, most of which are already near-zero, by assigning smaller weights to the detail subbands (keeping 5% of all the detail coefficients which account for 3/4 of the total). Notice also that our assignment of the weight 0.02 to the horizontal and vertical subbands and 0.01 to the diagonal subband comes from a reasonable assumption that the horizontal and vertical details are more common than the diagonal ones in natural images (Huang and Mumford, 1999; Watanabe et al., 1968). Thus, the weighted mean WSSIM can be interpreted as taking into account the approximation, horizontal detail, vertical detail and diagonal detail coefficients with appropriate weights as described above.

### 4.3 Wavelet-based non-parametric structure similarity index (WNPSSIM)

Luminance plays an important role when the human eye scans two images for similarity. We have found that WSSIM detached from luminance considerations assigns almost identical similarity indices for two noisy images that differ only in pixel-wise luminance shift (see  $N_{13,4}$  and  $N_{14,4}$  in Figure 9). But, the human eye, being sensitive to such change, prefers a better discrimination of their similarity indices. This objective is precisely met when we define our similarity index, called Wavelet-based Non-parametric Structure Similarity Index (WNPSSIM), to be the product of WSSIM and the average luminance similarity. Following Wang et al. (2011), we take the average luminance similarity  $L$  to be the average of pixel-wise luminance similarities instead of



the locally estimated mean pixel values as in SSIM to avoid the bias as a result of local smoothing for finite samples. For two images  $X = (X_{ij})$  and  $Y = (Y_{ij})$  of size  $m \times n$ , we thus have

$$\text{WNPSSIM}(X, Y) = p_{ap}^{\frac{19}{20}} \cdot p_h^{\frac{2}{100}} \cdot p_v^{\frac{2}{100}} \cdot p_d^{\frac{1}{100}} \cdot L,$$

where

$$L = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n L_{ij} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \frac{2X_{ij}Y_{ij} + C}{X_{ij}^2 + Y_{ij}^2 + C},$$

where, in turn,  $C$  is some small positive constant to avoid numerical instability which occurs when  $X_{ij}$  and  $Y_{ij}$  are very small floating-point numbers. In our experiment, we have used  $C = 0.001$ .

## 5. PERFORMANCE EVALUATION OF WNPSSIM

In this section, performance of WNPSSIM is evaluated against the performances of some well-established image similarity assessment methods in meeting the objective of this paper. The methods we have selected for this purpose are MSE, TV, SSIM, MS-SSIM, VIF, and IW-SSIM. MSE and its variant SNR (see Section 3.1) are traditional methods still widely used due to their mathematical convenience. TV is another mathematically convenient method which is more popular among the partial differential equations community. Just like in the case of MSE, TV's poor performance in trying to follow the HVS is its largest drawback which greatly subtracts from its recognition due to its mathematical convenience. SSIM (see Section 3.2), on the other hand, is highly regarded as the first successful universal image similarity assessment method in defeating traditional methods in capturing similarity in the sense of the HVS. Yet, SSIM does not fare very well in our experiment results. Our proposed method is specifically designed to perform well in the context of random-noise contamination and employs a multi-scale decomposition via the wavelet transform. MS-SSIM, VIF, and IW-SSIM (see Section 3.3) are all based on the techniques of a multi-scale decomposition of images. The MATLAB codes for SSIM and IW-SSIM were downloaded from <http://ece.uwaterloo.ca/~z70wang/research.htm> and the codes for MS-SSIM and VIF were downloaded from <http://live.ece.utexas.edu/research/Quality>. The WNPSSIM code, not publicly released yet but available from the authors upon request, was implemented in both MATLAB and R 2.8.1.

### 5.1 Experiment data

Our experiment data consist of nine standard images in the image processing literature as original reference images and 17 different noisy versions of each one of them as test images. For robustness, we include all common random noises

described in Section 2. The original reference images are labeled  $A_1, A_2, \dots, A_9$  (see Figure 8). All are of size  $512 \times 512$  and has intensities ranging from 0 to 255. Of each reference image  $A_i$ , 17 noisy versions were created by random noise contamination of various types and levels and labeled as  $N_{1,i}, N_{2,i}, \dots, N_{17,i}$ . These distortions are illustrated in Figure 9 for image  $A_4$ . The distorted images  $N_{13,i}$  and  $N_{14,i}$  were generated in R by adding mixture noise to the original images. The rest of the distortions were generated in MATLAB using the `imnoise.m` function. The numerical parameters of `imnoise.m` are normalized so that the images are scaled to have intensities in the range of  $[0, 1]$  before the noise contamination process and then scaled back to the original range of  $\{0, \dots, 255\}$  afterward. The description of these noisy images are given as below:

1.  $N_{1,i}, N_{2,i}, N_{3,i}, N_{4,i}$  and  $N_{5,i}$  contain additive Gaussian noise with mean 0 and variances 0.00068, 0.0018, 0.005, 0.01 and 0.068 respectively.
2.  $N_{6,i}, N_{7,i}$  and  $N_{8,i}$  contain salt and pepper noise with noise densities 0.0011, 0.006 and 0.011 respectively.
3.  $N_{9,i}$  contains Poisson noise generated from the image itself. If the input pixel value is  $\lambda$  then the output pixel value is drawn from the Poisson distribution with mean  $\lambda$ .
4.  $N_{10,i}, N_{11,i}$  and  $N_{12,i}$  contain MATLAB-generated speckle noise (addition of uniformly distributed multiplicative noise) with mean 0 and variances 0.007, 0.012 and 0.12 respectively.
5.  $N_{13,i}$  and  $N_{14,i}$  contain mixture noise coming from two sources: 60% of the noise is from an exponential distribution with mean 1 and 40% of the noise is from a t-distribution with 3 degrees of freedom with mean at 90 and 120 respectively. Note that the parameters here correspond to image pixel values in the range  $\{0, \dots, 255\}$ .
6.  $N_{15,i}, N_{16,i}$  and  $N_{17,i}$  contain localvar noise. This noise is generated in MATLAB as an additive Gaussian noise with mean 0 and variance as a function of image intensity. When at least two image intensities take two different variances, MATLAB extends this initial correspondence to be an all-out intensity-to-variance function by assigning a unique variance to each of the whole range of intensities in the particular image through linear interpolation. The initial intensity-to-variance correspondences used to generate images  $N_{15,i}, N_{16,i}$  and  $N_{17,i}$  are  $\{0.01, 0.09\} \mapsto \{0.02, 0.01\}$ ,  $\{0.01, 0.09, 0.9\} \mapsto \{0.08, 0.02, 0.01\}$  and  $\{0.01, 0.9\} \mapsto \{0.06, 0.1\}$  respectively.

Figure 4 plots the SNR values for random noisy versions of various types and levels of each of the 9 reference images. Figure 8 shows all 9 reference images and Figure 9 shows all 17 noises of one single reference image. The SNR values are only meant to show how much noise is present in our experimental data by using a traditional measure.

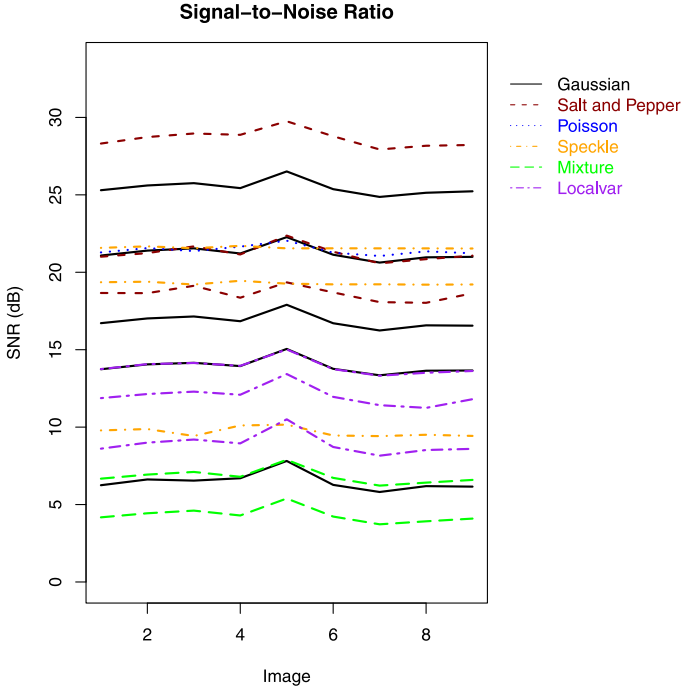


Figure 4. SNR values of 6 different types of random noisy versions of each of the 9 reference images. Full details of the noise types and levels are provided in Section 5.1. All the 9 reference images are illustrated in Figure 8 and all the 17 versions of random noise in a particular reference image is illustrated in Figure 9.

## 5.2 Comparison of WNPSSIM with other image similarity indices

In this section we provide experimental results to verify that WNPSSIM outperforms some popular image similarity methods in order to meet the main objective of this paper, that is, the ability to see through random noise in images and assign similarity indices that are more coherent with the human eye.

Six popular image similarity assessment methods, namely, MSE, TV, SSIM, MS-SSIM, VIF, and IW-SSIM are evaluated vis-à-vis WNPSSIM, the proposed method. However, keeping in mind space constraints and relevance to our current context as outlined in Section 3.3, we provide a complete list of similarity measures only for MS-SSIM, VIF, IW-SSIM and the five versions of WNPSSIM and make a passing reference to the similarity measures of MSE, TV and SSIM in our discussion. Despite the noises, all the images are clearly recognizable as noise-contaminated versions of the original image. MSE and TV are bound to perform the worst as they are meant to capture mathematical similarities rather than visual similarities between the two images. SSIM is also seen to assign values rather inconsistently and unreasonably. In order to save some space, we do not provide the SSIM values for our entire experiment data but

would like to point out that the SSIM values do appear to be changing too dramatically in the case of mixture, Gaussian and localvar noises in comparison to the effective visual differences.

Figure 5 shows the image similarity measures for all 17 noisy versions of each of the 9 reference images assigned by similarity indices MS-SSIM, VIF, IW-SSIM, and the five wavelet versions of WNPSSIM (denoted as HaarWNPSSIM, Db4WNPSSIM, Sym4WNPSSIM, Jpeg9.7WNPSSIM and Bior6.8WNPSSIM). Although we have five different versions of WNPSSIM associated to five different wavelets, we will constantly refer to HaarWNPSSIM as the representative WNPSSIM index since the five versions have similar performance.

Figures 1 and 6 show more details for comparison of HaarWNPSSIM with MS-SSIM, VIF, IW-SSIM. Figure 1 presents the similarity assessments as judged by these image similarity indices for the noisy versions of all the 9 reference images collectively on the left and the Box-Percentile plots for the same on the right. Figure 6 illustrates the performances of these similarity indices individually for each of the 9 images. For subjective evaluation of the random noise assessment, the reader can refer to Figure 9 which lists  $A_4$  as the original reference image as well as all the 17 noisy versions of it. The images in the paper are small for organizational purposes but they can be viewed in a larger scale by zooming in on the pdf file to see the degradation of the visual quality with higher levels of noise. For these images, the reader can compare their own subjective judgement of the visual similarity with the assessment of various similarity indices given in Figure 6.

First of all, notice that all the similarity measures in Figures 5 and 6 lie between 0 and 1, where the index value of 1 indicates perfect similarity (see Section 3.4). Another common feature about MS-SSIM, VIF, IW-SSIM and WNPSSIM is that all these indices agree on the relative rankings of the similarity measures for most of the different versions of noise contamination as illustrated in Figure 1. The only exception is between the noisy versions  $N_{8,k}$  and  $N_{9,k}$  of the original image  $A_k$ , ( $k = 1, \dots, 9$ ). Note that  $N_{8,k}$  was created by sprinkling salt and pepper noise thereby affecting its visual similarity with  $A_4$  only partially whereas  $N_{9,k}$  contains Poisson noise which affects the entire image. This observation is also endorsed by the illustration in Figure 9 where  $N_{9,4}$  appears to be slightly more corrupted than  $N_{8,4}$ . Hence, HaarWNPSSIM's assignment of slightly higher index values for  $N_{8,k}$  than for  $N_{9,k}$  appears to be more in accord with our visual perception. On the other hand, MS-SSIM, VIF and IW-SSIM are observed to assign much lower values for  $N_{8,k}$  than for  $N_{9,k}$  in direct contrast to our visual sense.

Another observation is that the similarity indices assigned by SSIM, MS-SSIM, VIF, and IW-SSIM have big ranges over our experiment data (see Figure 1). Scanning the images in Figure 9, however, suggests that the incremental noise level within each noise type changes slightly.

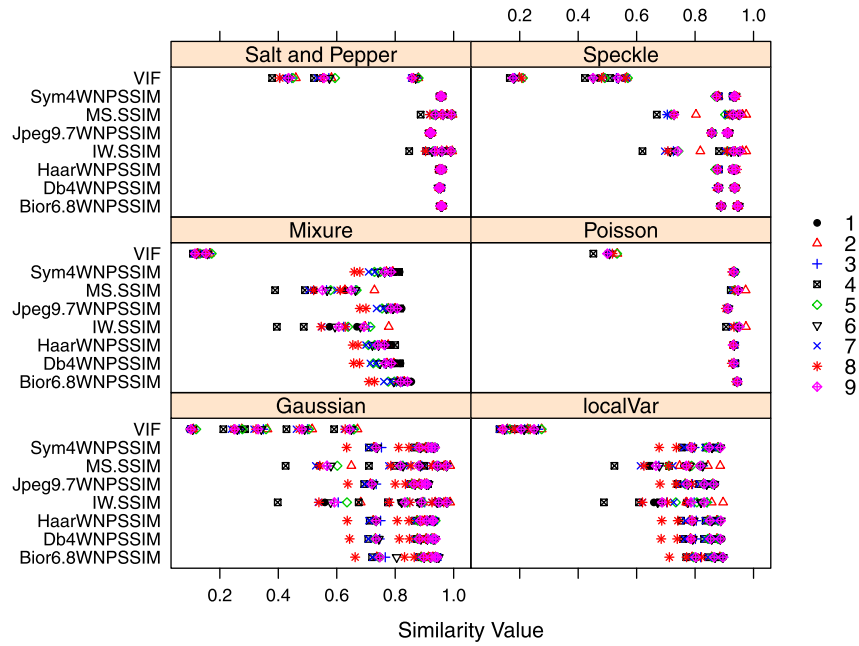


Figure 5. Visual illustration of the comparison of WNPSSIM with MS-SSIM, VIF and IW-SSIM for all 9 images and its 17 noisy versions of 6 different types. The original similarity values assigned by these methods are given in Tables 1–3 in supplementary material (<http://www.intlpress.com/SII/p/2013/6-1/SII-6-1-silwal-supplement.pdf>). Full details of the noise types and levels are provided in Section 5.1. All the 9 reference images are illustrated in Figure 8 and all the 17 versions of random noise in a particular reference image is illustrated in Figure 9. The five versions of WNPSSIM have very close performance. VIF assigns much lower similarity values than other methods. MS-SSIM and IW-SSIM show big variations.

The slight change should not be the cause for drastic changes in similarity index values. WNPSSIM assigns values within a shorter range ( $[0.713, 0.96]$  for HaarWNPSSIM), remaining more in line with the visual perception, while values assigned by the other indices fluctuate quite dramatically (for instance,  $[0.425, 0.988]$  for MS-SSIM and  $[0.101, 0.859]$  for VIF).

One attractive feature of WNPSSIM which makes it very convenient as an image similarity index is the fact that it is a symmetric index and rates similarity strictly on a scale of 0 to 1. MS-SSIM, VIF and IW-SSIM might not have these properties in general. Although MS-SSIM is a symmetric index, its values, in full generality, range between  $-1$  and  $1$ . VIF is an unsymmetric index and takes values bigger than 1 if the test image is an enhanced version of the original image. IW-SSIM is an asymmetric index and might take complex values. In our experiment data, the values assigned by MS-SSIM, VIF and IW-SSIM do lie between 0 and 1. However, these similarity index values cannot be interpreted on a scale of 0 to 1 as they clearly do not vary in accordance to perceived levels of random noise. The values assigned by these three similarity indices appear to plunge dramatically as the random noise contamination increases. The noisy versions  $N_{5,k}$ ,  $N_{8,k}$ ,  $N_{12,k}$ ,  $N_{14,k}$ , and  $N_{17,k}$  contain the highest level of Gaussian, salt & pepper, speckle, mixture, and localvar noises respectively. MS-SSIM, VIF, and IWSSIM assigned

very low similarity measures for these noisy versions. For example, for  $N_{14,4}$ , the similarity indices assigned by MS-SSIM, VIF, and IWSSIM are 0.3889, 0.1081, and 0.3952 respectively. Due to the complicated range for these similarity indices, it is difficult for the human brain to quickly judge how far away these values are from perfect similarity. The WNPSSIM value of 0.7726 in the range of  $[0, 1]$  for  $N_{14,4}$  better reflects our visual sense because although it contains the highest level of random noise coming from two different sources, a very high degree of similarity with the original image  $A_4$  is clearly visible to the human eye. Similarly, the similarity indices of 0.5243, 0.1311, and 0.4878 assigned respectively by MS-SSIM, VIF, and IW-SSIM for the noisy image  $N_{17,4}$  may be difficult for the user to judge how close this image is to the original reference image. The WNPSSIM value of 0.7516 strictly lying in the range of  $[0, 1]$  clearly indicates that the noisy image  $N_{17,4}$  exhibits about 75% resemblance of the original image.

An outstanding observation we have made about the performance of WNPSSIM is that it assigns more consistent values for each noise contamination across all 9 images over MS-SSIM, VIF, and IW-SSIM. It is clearly observed by inspecting the plot of the similarity measures for each of the images in Figure 1 or 6. This suggests that HaarWNPSSIM’s measure of visible degradation in images is noise-specific and robust to variations of images while the measures de-

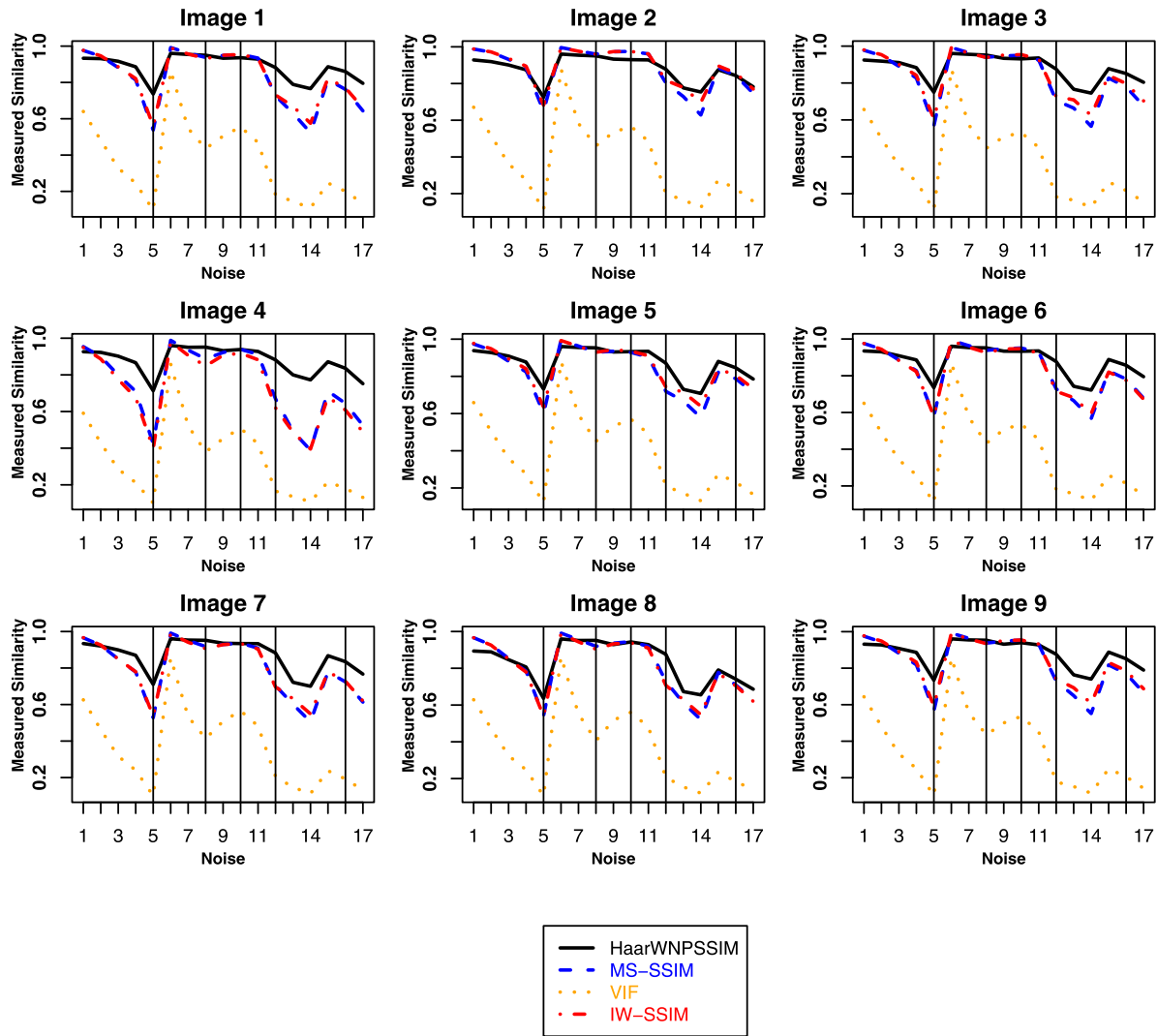


Figure 6. Visual illustration of the comparison of MS-SSIM, VIF, IW-SSIM and HaarWNPSSIM for similarity assessment of each of the 9 reference images with their 17 noisy versions. Full details of the noise types and levels are provided in Section 5.1. All the 9 reference images are illustrated in Figure 8 and all the 17 versions of random noise in a particular reference image is illustrated in Figure 9.

terminated by MS-SSIM, VIF, and IW-SSIM may be adversely affected by image-specific details present in images. An ideal image similarity index in presence of random noises should see through the noises in images and focus on the scene and features that have substantial effect on the HVS. Our experimentation results suggest that in comparison to MS-SSIM, VIF, and IW-SSIM, WNPSSIM indeed is more in agreement with this property of the HVS as it appears to be more consistent in perceiving the features that matter more to the human eye. Also, the Box-Percentile plot illustrated in Figure 1 reveals that, relative to the variations of measures assigned to different versions of noise contamination, the difference between MS-SSIM and IW-SSIM values are negligible. The VIF index is seen to have the largest

range and the lowest similarity index values for each type of noise. On the other hand, WNPSSIM has the lowest range and highest similarity index values in general in each type of noise. Our analysis shows that VIF would be more desired for detecting the slightest level of noise in images while WNPSSIM would be the best performer when it comes to detecting visual similarity in the presence of random noise.

In order to verify that the proposed similarity measure is also capable of detecting dissimilarity, we present similarity indices experimented on some completely different images. We first take the  $A_4$  (lady with black hair) as the reference image and compare it to  $A_8$ ,  $N_{2,1}$ ,  $N_{7,2}$ ,  $N_{14,8}$  and  $N_{14,4}$ . The results are presented in the first five rows underneath the images in Figure 2. Then we take the  $N_{14,4}$  as the refer-



ence image and compare it to  $A_4$ ,  $A_8$ ,  $N_{2,1}$ ,  $N_{7,2}$  and  $N_{14,8}$ . The results are shown in the last five rows in Figure 2. It can be seen that MS-SSIM, IW-SSIM, and WNPSSIM all detect dissimilarity reasonably well. Evidently, VIF is seen to be having difficulty detecting dissimilarity between two images. Taking  $N_{14,4}$  (lady with the highest mixture noise) as the reference image, the VIF index value for  $N_{14,8}$  (peppers with similar mixture noise) is 0.4396 which is much higher than the VIF index is 0.0223 for  $A_4$  (noise-free version of  $N_{14,4}$  itself). The HaarWNPSSIM index values for these comparisons are 0.1489 for the different image and 0.7726 for the same image source. Clearly, HaarWNPSSIM is following the eye whereas VIF is merely comparing the superficial noises rather than the underlying images, thus assigning a misleadingly higher similarity index value.

Finally, although WNPSSIM has values much higher than the other ones for these dissimilarity comparisons, it is to be noted that WNPSSIM is strictly on a scale of 0 to 1 while the others are not. The non-parametric test we have employed in the construction of our similarity index is essentially an “iid” noise detector. Hence, our similarity index is not expected to perform well in the case of some deterministic noises such as those caused by spatial filtering. In addition, a locally homogeneous region in one image compared to such a region in another image would give a non-rejection even though the two regions may represent different image contents or objects. Consequently, the cut-off threshold for unsimilarity for the entire images would shift upward away from 0. The general rule of thumb is that if WNPSSIM is below 0.4, the images should be taken to have significantly low structural similarity. This can be seen in our comparisons of completely different images:  $A_4$  and  $N_{14,8}$ ,  $N_{14,4}$  and  $A_8$ , and  $N_{14,4}$  and  $N_{7,2}$  in Figure 2, where WNPSSIM values are all between 0.25 and 0.4.

## 6. APPLICATIONS AND LIMITATIONS OF WNPSSIM

In this section, the applications and limitations of the proposed image similarity index, namely, WNPSSIM are described. We also discuss the question of whether or not WNPSSIM defines an actual distance.

### 6.1 Applications and limitations in image analysis

Algorithmic methods of image similarity assessment consistent with the HVS are most vigorously pursued research topics in the image analysis literature. Robust methods, whether universal or context-specific, will evidently have huge applications in several fields as pointed out in Section 1. In this article, a new image similarity assessment method by the name of WNPSSIM has been proposed in the context of images contaminated with random noises from all possible sources. Images are susceptible to random noise all too often such as during acquisition, storage and processing.

Our experiments show that WNPSSIM is quite powerful in detecting visual similarity between images when the images have been contaminated with random noises from various sources.

The construction of WNPSSIM we have described in this article worked out best for the experiments we have conducted which consisted of 9 standard  $512 \times 512$  test images common in the literature. Some parameters in the construction of WNPSSIM can be adjusted in order to better suit the application data as follows:

1. *Local window size*: At typical viewing distances, only a local area in the image can be perceived with high resolution by the human observer at one time instance because of the foveation feature of the HVS (Wang et al., 2004). For this reason, computation of local statistics is necessary. For  $512 \times 512$  images in this article, we have used a local window of size  $11 \times 11$  as in Wang et al. (2004) (see Section 4.2). With this choice there are enough pixels within each local window to make the test valid. We comment that the test is also valid if the local window size is bigger, which also reduces the computational requirement since there are fewer windows to be tested. However, increasing the local window size will make the algorithm depart from incorporating the foveation feature of the HVS. Hence, we recommend to use local windows of size  $11 \times 11$  for all images. With this choice of local window, we have found that our similarity index is capable of detecting significant visual similarity even when the test images have some image information missing. This is shown in Figure 7 where both of the reference images are of size  $512 \times 512$ . In test image 1, the last 51 columns, i.e., 9.96% of the total pixels were subjected to information loss by turning them into grey pixels of some fixed intensity. In test image 2, a near square area of  $100 \times 101 = 10100$  pixels in the upper right corner were blackened out resulting in a loss of 3.85% image information. On top of the information loss, low levels of random noise were also imposed on the test images. Test image 1 was contaminated with a speckle noise with mean 0 and variance 0.002 while test image 2 was corrupted with a salt & pepper noise with density 0.005. In this particular experiment illustrated in Figure 7, WNPSSIM finds the information loss more dominating relative to the random noise contamination and assigns similarity measures for the two test images in the reverse order of the information loss percentage. The assignments from the other measures are just opposite.
2. *Wavelet type*: There are a vast number of wavelets available for implementation with the wavelet toolbox in MATLAB that vary in symmetry, continuity, support size, etc. We have used five wavelets, namely, Haar, Daubechies 4, Symmlet 4, Jpeg 9.7 and Biorthogonal 6.8 in our experiment which are available with DIPUM package (Gonzalez et al., 2009). We notice





Reference Image 1	Test Image 1	
		
		WNPSSIM Haar = 0.7870 db4 = 0.7607 sym4 = 0.7935 jpeg9.7 = 0.7807 bior6.8 = 0.8002
	SNR=16.4016 SSIM = 0.9137 MS-SSIM = 0.9411 VIF = 0.6724 IW-SSIM = 0.9361	
Reference Image 2	Test Image 2	
		
		WNPSSIM Haar = 0.8535 db4 = 0.8774 sym4 = 0.8763 jpeg9.7 = 0.8487 bior6.8 = 0.8986
	SNR=12.1378 SSIM = 0.8691 MS-SSIM = 0.9344 VIF = 0.5552 IW-SSIM = 0.906	

Figure 7. The foveation feature of the HVS asserts that human observers only focus on one local area at a time. Illustration of this feature using images with lost partial information and low level random noise contamination. Test images were generated in MATLAB by partially removing or occluding some information from the corresponding reference images followed by corruption with random noise. Test image 1 has 51 columns turned to grey which amounts to 9.96% of information loss. Test image 2 has a subimage of size  $100 \times 101$  turned to black resulting in 3.85% information missing. Additionally, test image 1 contains speckle noise with mean 0 and variance 0.002; and test image 2 has salt & pepper noise with density 0.005.

slightly different results for these five different wavelets. This happens because wavelets come in different shapes and sizes and measuring an arbitrary signal in terms of different wavelets will produce different results. In fact, it is possible to find empirically a wavelet that works best for a given data set of images, such as faces, fingerprints, terrain, paintings, etc. For example, for a data set of aerial images of skyscrapers, Haar wavelets will provide the best succinct representation with least wavelet coefficients due to their block-like resemblance. However, we notice for our collection of experimental images, WNPSSIM based on the selected five wavelets produce similar measures and that they all perform better in comparison with the other image quality assessment methods considered in this article. If the applica-

tion data set is not well-defined, we recommend using Haar wavelets for their simplicity.

3. *Wavelet subband weights*: The weights we have used in this article for approximation, horizontal, vertical and diagonal subbands are 0.95, 0.02, 0.02 and 0.01 respectively (see Section 4.2). The reason for these choices come from our belief that for our experiment data, the visual structure of the original image retained by the approximation, horizontal, vertical and diagonal subbands are 95%, 2%, 2% and 1% respectively. Please refer to Figure 3 to have an estimate of these values. These parameters can be adjusted if there is a very strong reason to believe otherwise. However, we do not recommend to change these weights because it is very hard to ascertain the precise values of these parameters and also the decomposition of the original image into wavelet subbands fairly stays the same regardless of the image types.
4. *Wavelet decomposition level*: We have used only one level of wavelet decomposition in this article. Depending on the size of the images under consideration, higher levels of decomposition is possible. Even if more decomposition levels are possible, the authors recommend to use only one decomposition level as it keeps the algorithm fairly simple and it is very hard to justify the use of more iterations of the wavelet transform in our present context. Nevertheless, if more than one decomposition level is preferred, then the local window size and the wavelet subband weights need to be adjusted accordingly.

It should be noted that the application of WNPSSIM is to be restricted to registered databases of images. This means that all the images in the database need to be brought down to the same coordinate system and have the same size. In other words, image registration is a required pre-processing step if WNPSSIM is to be applied to an arbitrary set of images. This, in fact, is true of almost all currently available full-reference image quality assessment methods. We propose WNPSSIM as a robust method to detect image similarity very much in accordance with that of the HVS in the presence of random noise of all types. For that reason, we think image denoising might be the best application domain of WNPSSIM.

Finally, MS-SSIM, VIF and IW-SSIM have gone through validation involving extensive subjective experiments to follow the HVS in the same pattern. The fact that WNPSSIM concurs with MS-SSIM, VIF and IW-SSIM on relative rankings of the similarity measures for different kinds of noises with the exception of one case (see Section 5.2) we can conclude that WNPSSIM exhibits a reasonable amount of proximity with the HVS. Furthermore, WNPSSIM offers additional advantages (see Section 5.2) especially in the case of images with noises from multiple sources, namely,  $N_{13,k}$  and  $N_{14,k}$ , and images with noises having intensity-dependent variances, namely,  $N_{15,k}$  through  $N_{17,k}$ .

The authors are aware that there are some large-scale image databases featuring psycho-visual experimentation such as LIVE image quality assessment database (Sheikh et al., 2005b) and TID2008 database (Ponomarenko et al., 2009b) available for use completely free of charge to researchers. However, the main purpose in this paper is the detection of visual similarity between images degraded only through random noise contamination. Deterministic noises such as compression and blurring are beyond the scope of this paper. In order to establish the robustness of the proposed method, we need several different kinds of random noises at several different levels. Most publicly available databases, featuring both deterministic and random noises, are specially designed for universal image similarity assessment methods. However, due to an insufficient number of variations in random noise, they are not quite suitable to serve the specific purpose undertaken in this paper. This is essentially what prompted us to come up with our own experiment data. The authors were unable to locate any publicly available databases which offered the variety of random noises we have considered in this paper. For example, our experiment data features a mixture noise which is a random noise with more than one source. A mixture noise is certainly a realistic noise since images in real life subjected to various processes are susceptible to contamination from multiple sources. Our experiments show that this type of noise might adversely affect the performance of some popular image similarity indices such as SSIM. Also, the results in the context of the 5 relatively complicated noises  $N_{13,k}$  through  $N_{17,k}$  sufficiently convince the authors to refrain from resorting to any large-scale experiments. In light of the large proportion of time and effort such experimentation would require as well as its value in its own right, the authors wish to undertake it as a separate project in the future.

## 6.2 Does WNPSSIM define a distance?

It is relevant to raise a question as to whether or not WNPSSIM defines a distance. We give a negative answer to this question and provide an argument that a distance as an image similarity assessment method is not necessarily desirable. A distance is usually desired for its mathematical convenience due to the fact that it makes a continuous measurement of distances between images possible. A distance  $d$ , by definition, is a function taking values in  $[0, \infty)$  and possessing the following properties:

1.  $d(X, Y) = d(Y, X)$  for any two images  $X$  and  $Y$ .
2.  $d(X, Y) = 0$  if and only if the two images  $X$  and  $Y$  are identical.
3.  $d(X, Y) \leq d(X, Z) + d(Z, Y)$  for any three images  $X, Y$  and  $Z$ .

To a distance, an  $M \times N$  matrix is just a mathematical point in the space  $\mathbb{R}^{MN}$  and it treats all the points in this space exactly the same way without any discrimination. Consequently, a distance lacks the ability to take into account the

redundancy present in natural images as it cannot distinguish between random matrices and matrices corresponding to natural images. However, as we have remarked in Section 1, the whole purpose of image quality assessment is to provide a method that goes hand in hand with the HVS in determining the proximity between natural images. A distance in the space of all possible images (random and natural), clearly, will not do justice to the set of natural images which, according to the discussion in Section 4.1, is merely a thin subset of the space of all images and, therefore, cannot be the right tool for image quality assessment.

Given the range and meaning of WNPSSIM (i.e., the interval  $[0, 1]$ , with 0 indicating complete dissimilarity and 1 complete similarity), a natural candidate for a distance defined by WNPSSIM is  $d(X, Y) = -\log(\text{WNPSSIM}(X, Y))$ , which trivially verifies Properties 1 and 2. Along the lines of the previous discussion, the fact that  $d(X, Y)$  fails to satisfy Property 3 will attest to its ability to discriminate between natural images and mathematical matrices. This is indeed the case, for instance, if we consider  $X = A_4$ ,  $Y = A_8$  and  $Z = N_{14,4}$ , where the images  $A_4$ ,  $A_8$ , and  $N_{14,4}$  are as in Figure 2. Then we find from the WNPSSIM values in the same figure,

$$7.6009 = d(X, Y) \geq d(Z, Y) + d(X, Z) = 1.3556 + 0.2580.$$

Hence, as anticipated,  $d$  does not satisfy Property 3. In addition, the corresponding values of  $d(X, Y)$ ,  $d(Z, Y)$ , and  $d(X, Z)$  are in accordance with visual perception. Meaning, images  $X$  and  $Y$  are clearly different images,  $Z$  is a noisy version of  $X$  and our assessment of its distance to  $X$  and  $Y$  is hindered by the existence of noises leading to smaller distances.

## 7. CONCLUSION

One of the properties of the HVS is the ability to see through a low-level of random noise contamination and recognize the underlying image. Another important property of the HVS is its *multi-channel parallel pathways* (see Section 4.1). This article seeks to emulate these two properties in the proposed image similarity index.

In this article, we have developed an image similarity index WNPSSIM that provides a measure of visual similarity on a scale of 0 to 1 between any two supplied images. In practice, one of the images is a reference image and the other one is a random-noisy version of it and the goal is to assess the similarity between the two, incorporating the seeing-through ability of moderate levels of random noise in images. The main idea of the construction of WNPSSIM is to apply a non-parametric test to the relative difference of the reference image and the test image in the wavelet domain to evaluate the structural similarity of the images locally. The final similarity assessment method integrates the information from the level-one wavelet subbands as well as the local luminance comparison.





Figure 8. All 9 reference images with six different types of random noise.





Figure 9. Original image  $A_4$  and all of its 17 noisy versions.

Since the WNPSSIM index is a method based on multi-scale decomposition techniques, the three other multi-scale similarity indices, MS-SSIM, VIF, and IW-SSIM, have presented themselves as suitable choices for comparison against the performance of WNPSSIM in the present context of this article. Furthermore, these three similarity indices are extensively tested methods and are widely regarded as state-of-the-art performers in mimicking the human visual system (HVS). Our experimentation show that WNPSSIM indeed shares the virtues of MS-SSIM, VIF, IW-SSIM in detecting visual similarity between images very much in sync with the HVS. This is clear from the fact that MS-SSIM, VIF, IW-SSIM, and WNPSSIM almost always agree on the rankings of different types of noisy versions according to their visual similarity with the reference image. Our results find that WNPSSIM offers additional advantage over the other considered image similarity indices when it comes to random-noise contamination. WNPSSIM values comply more with our visual sense by assigning similarity values using the geometric mean of the proportions of local agreements between the two images in each wavelet subband scaled by the average luminance. Further, slight incremental changes in the levels of noise seem to affect the images in our experiment data only slightly in terms of visual degradation. This fact is reflected very well in WNPSSIM values. The other similarity indices, on the other hand, show drastic changes in

their values for these noisy images. Since various kinds of random noises inevitably exist in abundance in digital images, a non-parametric image quality assessment method as the one proposed in this article should have plenty of potential to provide a reliable similarity measure in this context.

## ACKNOWLEDGEMENTS

This work was partially supported by a grant from the National Science Foundation (DMS 0901587 to Diego Maldonado) and a grant from the Simons Foundation (#246077 to Haiyan Wang).

*Received 14 May 2012*

## REFERENCES

- BOVIK, A. (2009). *The Essential Guide to Image Processing*. Academic Press.
- CHANDLER, D. M. and HEMAMI, S. S. (2005). Dynamic contrast-based quantization for lossy wavelet image compression. *IEEE T. Image Process.* **14**(4) 397–410.
- CHANDLER, D. M. and HEMAMI, S. S. (2007). A wavelet-based visual signal-to-noise ratio for natural images. *IEEE T. Image Process.* **16**(9) 2284–2298. [MR2468097](#)
- DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics. [MR1162107](#)

- DONOHO, D. L. (1995). Denoising via wavelet shrinkage. *IEEE T. Inform. Theory* **41**(3) 613–627.
- DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**(3) 425–455. [MR1311089](#)
- EGIAZARIAN, K., ASTOLA, J., PONOMARENKO, N., LUKIN, V., BATTISTI, F., and CARLI, M. (2006). New full-reference quality metrics based on HVS. In: *CD-ROM Proc. of the Second Int. Workshop on Video Proc. and Quality Metrics*.
- FIELD, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am.* **4**(12) 2379–2394.
- FIELD, D. J. (1999). Wavelets, vision and the statistics of natural scenes. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences* **357**(1760) 2527–2542.
- FIELD, D. J. and BRADY, N. (December 1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Res.* **37**(23) 3367–3383.
- GAGNON, L. and JOUAN, A. (1997). Speckle filtering of SAR images – a comparative study between complex-wavelet based and standard filters. In: *Proc. SPIE* **3169** 80–91.
- GAO, X., LU, W., TAO, D., and LI, X. (2009). Image quality assessment based on multiscale geometric analysis. *IEEE T. Image Process.* **18**(7) 1409–1423. [MR2750639](#)
- GAO, X., LU, W., TAO, D., and LI, X. (2010). Image quality assessment and human visual system. In: *Proc. SPIE* **7744** 77440Z.
- GONZALEZ, R., WOODS, R., and EDDINS, S. (2009). *Digital Image Processing Using Matlab*. Gatesmark Publishing.
- HUANG, J. and MUMFORD, D. (1999). Statistics of natural images and models. *Computer Vision and Pattern Recognition* **1** 1541–1547. [MR2701355](#)
- KIM, M. H. and AKRITAS, M. G. (2010). Order thresholding. *Ann. Stat.* **38**(4) 2314–2350. [MR2676891](#)
- LAI, Y.-K. and KUO, C.-C. J. (2000). A Haar wavelet approach to compressed image quality assessment. *J. Vis. Commun. Image R.* **11** 17–40.
- LUISIER, F., VONESCH, C., BLU, T., and UNSER, M. (2010). Fast inter-scale wavelet denoising of poisson-corrupted images. *Signal Process.* **90**(2) 415–427.
- MALLAT, S. G. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3rd edition. [MR2479996](#)
- MALLAT, S. G. (July 1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE T. Pattern Anal.* **2**(7) 674–693.
- MILLANE, R. P., ALZAIDI, S., and HSIAO, W. H. (2003). Scaling and power spectra of natural images. *Image Vision Comput.* **14**(12) 2117–2128. Palmerston North, NZ.
- NINASSI, A., MEUR, O. L., CALLET, P. L., and BARBA, D. (2008). On the performance of human visual system based image quality assessment metric using wavelet domain. In: *Proc. SPIE Human Vision and Electronic Imaging XIII*, (6806).
- NINASSI, A., MEUR, O. L., CALLET, P. L., BARBA, D., and TIREL, A. (2006). Task impact on the visual attention in subjective image quality assessment. In: *Proc. SPIE 14th European Signal Process. Conf.* Florence, Italy.
- PONOMARENKO, N., BATTISTI, F., EGIASARIAN, K., ASTOLA, J., and LUKIN, V. (Jan 2009a). Metrics performance comparison for color image database. In: *4th Intl. Workshop Video Process. Quality Metrics Consumer Electron.* (6), p. 6.
- PONOMARENKO, N., LUKIN, V., ZELENKY, A., EGIASARIAN, K., CARLI, M., and BATTISTI, F. (2009b). TID2008 – a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics* **10**.
- PORTILLA, J., STRELA, V., WAINWRIGHT, M. J., and SIMONCELLI, E. P. (2003). Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE T. Image Process.* **12**(11) 1338–1351. [MR2026777](#)
- RICHTER, T. and LARABI, C. (2008). Toward objective image quality metrics: the AIC Eval Program of the JPEG. In: *Proc. of SPIE* **7073**(1) 70730X. doi:10.1117/12.795096.
- RUDERMAN, D. L. (1997). Origins of scaling in natural images. *Vision Res.* **37**(23) 3385–3398.
- SHEIKH, H. R. and BOVIK, A. C. (2006). Image information and visual quality. *IEEE T. Image Process.* **15**(2) 430–444.
- SHEIKH, H. R., BOVIK, A. C., and DE VECIANA, G. (2005a). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE T. Image Process.* **14**(12) 2117–2128.
- SHEIKH, H. R., WANG, Z., CORMACK, L., and BOVIK, A. C. (2005b). LIVE image quality assessment database, release 2, available at <http://live.ece.utexas.edu/research/quality>.
- SIMONCELLI, E. P. and ADELSON, E. H. (1996). Noise removal via bayesian wavelet coring. *IEEE T. Image Process.* **14**(12) 2117–2128.
- TOLHURST, D. J., TADMOR, Y., and CHAO, T. (1992). Amplitude spectra of natural images. *Ophthalm. Physiol. Opt.* **12**(2) 229–232.
- VONESCH, C., AGUET, F., VONESCH, J., and UNSER, M. (2006). The colored revolution of bioimaging. *IEEE Signal Process. Mag.* **3** 20–31.
- WAINWRIGHT, M. J., SIMONCELLI, E. P., and WILSKY, A. S. (2001). Random cascades on wavelet trees and their use in analyzing and modeling natural images. *Appl. Comput. Harmon. Anal.* **11** 89–123. [MR1841335](#)
- WANG, H. and AKRITAS, M. G. (2004). Rank tests for ANOVA with large number of factor levels. *J. Nonparametr. Stat.* **16**(3-4) 563–589. [MR2073042](#)
- WANG, H., MALDONADO, D., and SILWAL, S. (2011). A nonparametric-test-based structural similarity measure for digital images. *Comput. Stat. Data. An.* **55**(11) 2925–2936. [MR2813056](#)
- WANG, H., NEILL, J., and MILLER, F. (2008). Nonparametric clustering of functional data. *Statistics and its Interface* **1** 47–62. [MR2425344](#)
- WANG, H., TOLOS, S., and WANG, S. (2010). A distribution-free non-parametric test to detect dependence between a response variable and covariate in presence of heteroscedastic treatment effects. *Can. J. Stat.* **38**(3) 408–433. [MR2730117](#)
- WANG, Z. and BOVIK, A. (2002). A universal image quality index. *IEEE Signal Proc. Let.* **9**(3) 81–84.
- WANG, Z., BOVIK, A., SHEIKH, H., and SIMONCELLI, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE T. Image Process.* **13**(4) 600–612.
- WANG, Z. and BOVIK, A. C. (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal. Proc. Mag.* **26**(1) 98–117.
- WANG, Z. and LI, Q. (2011). Information content weighting for perceptual image quality assessment. *IEEE T. Image Process.* **20**(5) 1185–1198. [MR2838671](#)
- WANG, Z. and SIMONCELLI, E. P. (2005a). Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In: *Proc. of SPIE Human Vision and Electronic Imaging* **5666** 149–159.
- WANG, Z. and SIMONCELLI, E. P. (2005b). Translation insensitive image similarity in complex wavelet domain. In: *Proc. IEEE Inter. Conf. Acoustic, Speech and Signal Proc.* **2** 573–576.
- WANG, Z., SIMONCELLI, E. P., and BOVIK, A. C. (Nov. 2003). Multi-scale structural similarity for image quality assessment. In: *Proc. 37th IEEE Asilomar Conf. on Signals, Systems, and Computers*.
- WATANABE, A., MORI, T., NAGATA, S., and HIWATASHI, K. (1968). Spatial sine-wave responses of the human visual system. *Vision Res.* **8** 1245–1263.
- WATSON, A. B. (1987). The cortex transform: Rapid computation of simulated neural images. *Comput. Vis. Graph. Image Process.* **39** 311–327.
- XIONG, Z., RAMCHANDRAN, K., and ORCHARD, M. T. (1997). Space-frequency quantization for wavelet image coding. *IEEE T. Image Process.* **6** 677–693.

Sharad Silwal  
Northland College  
1411 Ellis Avenue  
Ashland, WI 54806  
USA  
E-mail address: [ssilwal@northland.edu](mailto:ssilwal@northland.edu)

Diego Maldonado  
Department of Mathematics  
Kansas State University  
Manhattan, KS 66506  
USA  
E-mail address: [dmaldona@math.ksu.edu](mailto:dmaldona@math.ksu.edu)

Haiyan Wang  
Department of Statistics  
Kansas State University  
Manhattan, KS 66506  
USA  
E-mail address: [hwang@ksu.edu](mailto:hwang@ksu.edu)