# DenoiseLab Philosophy: A Standard Test Set and Evaluation Method to Compare Denoising Algorithms

Steven Lansel
slansel@stanford.edu
School of Electrical Engineering
Stanford University

October 2, 2007

**Abstract**

Denoising is the task of attempting to remove unwanted noise from a corrupted signal in order to recover the original signal. We propose a set of original noise-free images and 195 noisy images along with an evaluation procedure to be used as a standard to compare the performance of denoising algorithms. The 13 original images were corrupted with additive white Gaussian noise (AWGN), multiplicative white Gaussian noise (MWGN), and Poisson noise at five different noise levels in order to create a standard set of 195 noisy images. For evaluation, each denoising algorithm generates denoised images from this standard test set of noisy images. Algorithms are evaluated upon traditional mean squared error (MSE) and peak signal to noise ratio (PSNR) of the denoised images as well as the Structural SIMilarity (SSIM) index, which provides a measurement of the perceptual visual quality of the images. The standard data set and evaluation method are provided by a MATLAB program called DENOISELAB. The program, documentation, noisy images, and algorithm evaluation data for comparison can be downloaded from `http://www.stanford.edu/~slansel/DenoiseLab`.

# 1 Introduction

Denoising of images is an important problem with real-world applications. Throughout the last 15 years (and beyond), a large number of algorithms have been proposed for image denoising. Often authors compare their proposed algorithm to recent state-of-the-art denoising algorithms and claim comparable or superior performance citing a small number of test images that possibly were chosen to highlight the advantages of the particular algorithm. With such a large number of algorithms claiming state-of-the-art performance, it is very difficult to understand how the performance of algorithms will compare on a particular denoising task. This paper proposes a standard data set and performance evaluation method for comparison of denoising algorithms, which is available in DenoiseLab, a Matlab software package that is publicly available online at `http://www.stanford.edu/~slansel/DenoiseLab`. DenoiseLab is capable of automatically generating the evaluations from denoised images, loading results from other algorithms, and plotting the evaluation metrics as desired by a user of the DenoiseLab graphical user interface.

The hope of the standard outlined in this paper is that it will be adopted by the denoising community and will:

- Aid in comparing algorithm performance on a number of standard test images containing both AWGN and signal-dependent noise distributions.

- Offer a simple standard comparison of algorithms by averaging over the 13 original noise-free images to obtain evaluation measurements of an algorithm for a 'typical image.'

- Encourage the community to evaluate algorithm performance based upon perceptual quality measures in addition to traditional error measures.

- Inspire future research and improvements by identifying strengths and weaknesses of algorithms.

- Enable authors to quickly and easily perform the evaluation and generate performance plots using DenoiseLab.

We hope DenoiseLab will have the same effect on the denoising community as other standard data sets and evaluation procedures have had on other research communities. For example, facial recognition research benefitted greatly from the FERET (Face Recognition Technology) standard [7].

Recently, the PSNR of denoised images from the test images from [1] have been used to compare denoising algorithms. The following papers contain results that are similar to the above evaluation method but may contain some different noisy images [2], [3], [4], [5], [6]. This shows that the community is ready for a standard denoising data set and evaluation method. Our proposed standard contains many of the same original noise-free images as [1] but also additional original images and a standard evaluation procedure that is more comprehensive than the PSNR. We have also chosen all of our images to have a standard size, 512x512, in order to permit averaging over the images to obtain a single overall performance evaluation instead of one for each of the original images.

The remainder of the paper is organized as follows. Section 2 describes the selection of the original noise-free images. Section 3 outlines the derivation and motivation of the noisy image. Exactly what information a denoising algorithm is granted access to is outlined in Section 4. Section 5 introduces and motivates the perceptual quality measurement, SSIM. Section 6 concludes with the advantages and disadvantages of the proposed standard.

# 2 Original Noise-Free Images

All images in DenoiseLab are 8 bit grayscale images with 512x512 pixels. The original images with the exception of `fingerprint` are natural images in the sense that they are natural scenes instead of texture images. All of the original images selected are very commonly used in the image processing community. The names of the original images are: `airplane`[3] (aka F-16), `barbara`[3,4,5], `boat`[1,4,5], `couple`[4], `elaine`[1,3], `fingerprint`[4,5], `goldhill`[2,3] (aka hill), `lena`[4], `man`[1], `mandrill` [2,3] (aka baboon), `peppers`[2,3], `stream`[1,3] (aka bridge), `zelda`[2,3].

The above references give the source(s) of the image in the exact same form except the `man` image, which was reduced in size from a 1024x1024 image by taking the average over 4 adjacent pixels. All of these images are relatively clear and noise-free. This is important because the original image is assumed to be of perfect quality and for evaluation denoising algorithms should not be penalized for removing noise inherent in the original image.

Unfortunately, each image often has several different versions that are commonly used for research. Variations can result from differences in cropping, zooming, and converting from color to gray scale, and numerical errors when converting or saving the image again. For example, the `barbara`[2] and `boat`[2,3] images have two commonly used versions that differ in how they were obtained from cropping a larger image. Since different versions of an image will behave differently for denoising tasks, it is difficult to compare denoising algorithm results on an image unless an author verifies that the original images were identical, which can take considerable effort. For this reason, we hope the denoising community can adopt these images and these versions in particular as a standard data set.

# 3 Noisy Images

DenoiseLab provides a set of 195 standard noisy images derived from the 13 original noise-free images. Each original image is corrupted with noise from three different noise distributions using 5 different noise levels.

Although additive white Gaussian noise (AWGN) is often assumed to be the only corruption mechanism in many denoising publications, this may not be the case in real-world denoising applications [8]. Depending on the imaging device and environmental conditions such as lighting level, the AWGN model may poorly reflect the actual noise in the image and the noise may instead be signal-dependent. Denoising algorithms designed for signal-independent noise may perform poorly on images corrupted with signal-dependent noise. It is interesting to test denoising algorithms on non-AWGN noise models in order to see how successful they are at general denoising and if they are very sensitive to the assumption of noise perfectly modeled by AWGN. Additionally, DenoiseLab allows algorithms that are specifically designed for signal-dependent noise to take advantage of knowledge of the particular noise distribution that was used to corrupt an image.

For this section, let $x(i)$ and $y(i)$ be the pixels (ordered lexicographically)

[1] `http://sipi.usc.edu/services/database/database.cgi?volume=misc`

[2] `http://sampl.ece.ohio-state.edu/database.htm`

[3] `http://decsai.ugr.es/cvg/dbimagenes/g512.php`

[4] `http://decsai.ugr.es/~javier/denoise/test_images/`

[5] `http://www.cs.tut.fi/~foi/GCF-BM3D/index.html#ref_software`

at location $i \in \mathbb{Z}$, $1 \leq i \leq N$ of the original noise-free and noisy images, respectively, where $N$ is the number of pixels in an image. Note that the standard test images were generated by rounding $y(i)$ to the nearest integer with clipping if necessary. Let $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x(i)$ and $\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x(i)-\overline{x})^2$. All of the noisy images were generated in a manner so that the following hold

$$E[x(i) - y(i)] = 0 \text{ for all } i \tag{1}$$

$$E[(x(i) - y(i))^2] = \sigma^2 \tag{2}$$

where the expectation in (2 is taken over all $i$. Therefore, any image generated using the same value of $\sigma$ and a different noise distribution are in a sense 'equally-noisy.' For each noise distribution, noisy images were generated using $\sigma \in \{5, 10, 15, 20, 25\}$ for each original noise-free image. The $\sigma$ values were chosen so that the noise spans the range from nearly imperceptible to very strong. Descriptions of the statistics of the noisy images and motivation for each noise distribution are provided below.

## Additive White Gaussian Noise (AWGN)

For AWGN, the value of a pixel in the noisy image is given by $y(i) = x(i) + n(i)$ where $n(i)$ is normal with mean 0 and variance $\sigma^2$ and $n(i)$ is independent from $x(j)$ and $n(k)$ for $\forall j, k$ where $k \neq i$. Often noise caused from the electrical components in an imaging device during image capture or transmission are modeled as AWGN.

## Multiplicative White Gaussian Noise (MWGN)

For MWGN, the value of a pixel in the noisy image is given by $y(i) = x(i)n(i)$ where $n(i)$ is normal with mean 1 and variance $\sigma_{MWGN}^2 = \frac{\sigma^2}{\sigma_x^2 + \overline{x}^2}$ and $n(i)$ is independent from $x(j)$ and $n(k)$ for $\forall j, k$ where $k \neq i$. The mean and variance of $n(i)$ were chosen in order to satisfy (1) and (2).

Multiplicative noise is common in many real-world images, and a number of denoising algorithms have been created to deal with multiplicative noise [9],[10],[11],[12]. Speckle noise is common for images derived from a coherent imaging device such as radar, laser, sonar, or ultrasound. A multiplicative noise model is most commonly used for speckle noise with the multiplicative distribution being Gaussian, exponential, or Rayleigh depending on the specific imaging device [8], [13].

The noise is signal-dependent for the MWGN case. This causes the noise in the brighter areas of an image to have a larger variance and appear more noisy.

### Poisson Noise

For Poisson noise, the value of a pixel in the noisy image is given by $y(i) = \frac{\dot{y}(i)}{\lambda}$ where

$$P(\dot{y}(i)) = \frac{(\lambda x(i))^{\dot{y}(i)} \exp(-\lambda x(i))}{\dot{y}(i)!}$$

$\lambda = \frac{\bar{x}}{\sigma^2}$ and $y(i)$ and $x(j)$ are independent for $i \neq j$. The normalization from $\dot{y}(i)$ to $y(i)$ and selection of $\lambda$ are necessary in order to satisfy (1) and (2). [8]

Poisson noise is common for images obtained in low-light conditions where the arrival of photons is modeled by a shot noise following a Poisson distribution. Examples of Poisson denoising methods can be found in [8] and [14].

## 4 Information Available to Denoising Algorithms

Although in real-world applications no knowledge about the original noise-free image other than the noisy image realization is available to the denoiser, in denoising experiments some knowledge is typically assumed about the noisy distribution. This is justified by the fact that knowledge about the imaging device and environment can determine which noise model to use and heuristics exist to estimate the relevant constants for a noise distribution from a noisy image. For the AWGN case, the actual value of $\sigma$ is typically known to the denoiser, which we allow in the proposed standard in addition to always allowing the denoiser to know whether the noisy image was generated from the AWGN, MWGN, or Poisson models.

For the MWGN and Poisson noise distributions, the corresponding constants that define the distribution of the noisy images given the noise-free images are $\sigma_{MWGN}$ and $\lambda$, respectively. Since many denoising algorithms assume an AWGN model, they require the corresponding $\sigma$ value in order to denoise the MWGN and Poisson noisy images. Other denoising algorithms that do not necessarily assume the AWGN model may prefer to have knowledge of the value of $\sigma_{MWGN}$ or $\lambda$ instead of the corresponding value of $\sigma$.

Although the proposed standard could allow denoising algorithms to have knowledge of both $\sigma$ and $\sigma_{MWGN}$ for MWGN noisy images or $\sigma$ and $\lambda$ for Poisson noisy images, this would not only reveal knowledge of the corrupting noise distribution but also knowledge of values describing the original noise-free distribution. For example, $\lambda * \sigma^2$ gives the overall mean of the original noise-free image. For this reason, we allow denoising algorithms to take advantage of knowledge of only one of the above parameters at a time. Specifically, for MWGN noisy images the denoiser can use the value of either $\sigma$ or $\sigma_{MWGN}$, but not both. Similarly for Poisson noisy images and $\sigma$ or $\lambda$. Additionally, algorithms that choose to use knowledge of the value of $\sigma_{MWGN}$ or $\lambda$ cannot attempt to discern the value of $\sigma$ by using the fact that $\sigma$ is a multiple of 5 since this is not feasible in practice.

Denoising algorithms are not allowed to use any knowledge from the original noise-free images such as for training an algorithm. The original noise-free images are strictly for testing and no direct knowledge from them should be used.

# 5 Perceptual Quality Evaluation

## Motivation

Many authors of papers presenting denoising algorithms have compared the performance of their algorithm to other algorithms based almost entirely upon some $L_2$ based norm such as mean squared error (MSE) or peak signal to noise ratio (PSNR). Although error measures such as MSE and PSNR are common in the denoising community, are simple and easy to compute, and may lead to easily solvable optimization problems, they are not reflective of the quality of images as perceived by the human visual system. For example, adding a constant amount to all pixels in an image or scaling the contrast of an image will result in images that have a high MSE (low PSNR) but will seem nearly identical to the original picture as perceived by a human observer.

Due to the poor performance of the MSE and PSNR for images that are to be viewed by humans, authors perhaps also make some vague unsubstantiated subjective claim that the visual quality of their denoised images is superior to denoised images from other algorithms or that artifacts from their algorithm are less disturbing. The community needs some standard objective and quantitative way of assessing the perceptual quality of denoised images. Ultimately the best way to judge the perceptual quality of images is by conducting an experiment where actual human subjects judge images such as in [15]. Unfortunately such studies are prohibitively expensive, time-consuming, and potentially dependent on the experimental setup such as the lighting, viewing distance, and angle. The denoising community would benefit greatly from an objective standard perceptual quality measurement that can be calculated easily and can predict reasonably accurately how well humans will judge the quality of denoised images.

## SSIM Description

We propose the Structural SIMilarity (SSIM) index as the perceptual quality measurement for denoised images [16]. Let $\mathbf{X}$ be an image, $x_i$ be the pixel at location $i$, and $\mathbf{x}_j$ be the set of pixels in a 11x11 window centered at location $j$. Let the above notation hold for a second image, $\mathbf{Y}$. The SSIM is a measurement of the similarity between two 11x11 image neighborhoods. The neighborhoods are weighted by a circular-symmetric Gaussian with standard deviation 1.5 that is normalized to have a unit sum of 1. Let the weights of the Gaussian distribution centered at location $j$ be given by $w_{j,i}$ for $i$ in the 11x11 window centered at $j$. The estimates of the local statistics of the image neighborhoods, $\mathbf{x}_j$ and $\mathbf{y}_j$, required for the SSIM are given by:

$$\begin{aligned}
\mu_{\mathbf{x}_j} &= \sum_{x_i \in \mathbf{x}_j} w_{j,i} x_i \\
\sigma_{\mathbf{x}_j} &= \left( \sum_{x_i \in \mathbf{x}_j} w_{j,i} (x_i - \mu_{x_j})^2 \right)^{\frac{1}{2}} \\
\sigma_{\mathbf{xy}_j} &= \sum_{\substack{x_i \in \mathbf{x}_j \\ y_i \in \mathbf{y}_j}} w_{j,i} (x_i - \mu_{x_j})(y_i - \mu_{y_j})
\end{aligned}$$

and similarily for $\mu_{\mathbf{y}_j}$ and $\sigma_{\mathbf{y}_j}$. Then, the SSIM between the two local image neighborhoods is defined as

$$SSIM(\mathbf{x}_j, \mathbf{y}_j) = l(\mathbf{x}_j, \mathbf{y}_j) c(\mathbf{x}_j, \mathbf{y}_j) s(\mathbf{x}_j, \mathbf{y}_j)$$

where

$$
\begin{aligned}
l(\mathbf{x}_j, \mathbf{y}_j) &= \frac{2\mu_{x_j}\mu_{y_j} + C_1}{\mu_{x_j}^2 + \mu_{y_j}^2 + C_1} \\
c(\mathbf{x}_j, \mathbf{y}_j) &= \frac{2\sigma_{x_j}\sigma_{y_j} + C_2}{\sigma_{x_j}^2 + \sigma_{y_j}^2 + C_2} \\
s(\mathbf{x}_j, \mathbf{y}_j) &= \frac{\sigma_{xy_j} + C_3}{\sigma_{x_j}\sigma_{y_j} + C_3}
\end{aligned}
\tag{3}
$$

are the luminance, contrast, and structural components, respectively. The constants $C_1$, $C_2$, and $C_3$ are included to avoid numerical instabilities in the ratios. The luminance, contrast, and structural components penalize the image neighborhoods for having different local means, variances, and joint statistics, respectively.

The SSIM, luminance, contrast, and structural components all are symmetric, nonnegative, and bounded above by 1. The SSIM also has the property of a unique maximum where $SSIM(\mathbf{x_j}, \mathbf{y_j}) = 1$ if and only if $\mathbf{x_j} = \mathbf{y_j}$.

Since an overall measure of perceptual similarity between images is desired, the following definitions extend the above measures from small neighborhoods to entire images:

$$
\begin{aligned}
MSSIM(X,Y) &= \frac{1}{M}\sum_{j=1}^{M} SSIM(\mathbf{x}_j, \mathbf{y}_j) \\
MLuminance(X,Y) &= \frac{1}{M}\sum_{j=1}^{M} l(\mathbf{x}_j, \mathbf{y}_j) \\
MContrast(X,Y) &= \frac{1}{M}\sum_{j=1}^{M} c(\mathbf{x}_j, \mathbf{y}_j) \\
MStructure(X,Y) &= \frac{1}{M}\sum_{j=1}^{M} s(\mathbf{x}_j, \mathbf{y}_j)
\end{aligned}
\tag{4}
$$

where $M = (512 - 10)^2$ is the number of local neighborhoods in 512x512 images.

Since the goal of denoising is to generate a denoised image that is as close as possible to the original image, the similarity measurements above are used with one image, say $X$, being the original noise-free image and the other image, say $Y$, being a denoised image. In this manner, the SSIM index acts as a full-reference quality assessment method. Therefore, the $MSSIM$ gives a measurement of how close the denoised image is to the original image perceptually. The $MLuminance$, $MContrast$, and $MStructure$ values give measurements of how well the denoised image matches the original image with respect to the local luminance, contrast, and correlation (structure), respectively. In addition, the above measurements are all between 0 and 1 with a score of 1 being given only if the denoised image is exactly equivalent to the original image.

DenoiseLab uses $C1 = (0.01 * 255)^2$, $C2 = (0.03 * 255)^2$, and $C3 = C2/2$ as recommended in [16]. DenoiseLab uses the software available at `http://www.cns.nyu.edu/~lcv/ssim/` with only slight modification.

## Reasons for Selecting SSIM

The following are reasons why the SSIM index was chosen as the best method for perceptual quality measurement for denoising:

- The SSIM index has a high correlation with the mean opinion score (MOS), which is determined experimentally with human observers. The SSIM outperforms other state-of-the-art objective measures on a number of measures of correspondence with MOS. The SSIM index also has motivation in an understanding of the human visual system (HVS) [16].
- The SSIM index gives a perceptual score between 0 and 1 to all images, which gives a comparison to the upper limit of denoising and permits averaging over multiple images to give an overall evaluation score for typical images.
- In addition to the MSSIM value for a denoised image, the MLuminance, MContrast, and MStructure values can be calculated to give a standard measure of the local luminance, contrast, and structural quality of a denoised image. These quantities can identify the relative strengths and weaknesses of denoising algorithms.

- The SSIM index produces perceptual quality images (each pixel is given a value based on how well the neighborhood centered at that pixel matches the original). In addition, similar images giving the local luminance, contrast, and structure components can be generated. These images can help determine where algorithms perform well or poorly in a particular image.

- The SSIM index is based on a simple formula that only compares small neighborhoods between two images unlike many bottom-up approaches that depend on empirical measurements and global statistics of the image. The simplicity can be incorporated into the design of new denoising techniques that are (nearly) SSIM optimal.

- There are no underlying assumptions about image statistics in the SSIM index that could favor a particular approach to denoising. For instance, the perceptual quality measure proposed in [17] and the denoising algorithm proposed in [1] assume natural scene statistics follow a Gaussian scale mixture in the wavelet domain.

The recent study [15] that was based on responses from human observers assessed the perceptual quality of some state-of-the-art denoising schemes. The 8 bit 512x512 `barbara`, `goldhill`, and `face` images with noise from an AWGN model with $\sigma = 15$ were used. Although currently only three of the algorithms from the study have been evaluated in DenoiseLab, the relative ranking of these top algorithms with the SSIM index coincides perfectly for the `barbara` and `goldhill` images. Once the other algorithms in the study are evaluated in DenoiseLab, we will be able to make a stronger comparison between the SSIM index and this study.

# 6 Conclusion

The standard data set and evaluation method proposed here has the potential to assist the denoising community by providing a generally accepted method of comparing algorithm performance. The standard encourages the evaluation of algorithms over a large number of original noise-free images and three different noise distributions. The proposed perceptual performance measurement, the SSIM, offers a number of advantages over traditional MSE and PSNR error measurements. In addition, the SSIM and standard as a whole enable researchers to identify images and image features where an algorithm performs well or poorly. DenoiseLab provides a convenient way to perform the evaluation and generate plots.

There are some valid criticisms of the standard presented here. Algorithm performance is only evaluated on images that are 512x512 pixels. It is possible that the relative performance of algorithms will change for smaller or larger images. Since only 13 original noise-free images are used, perhaps the evaluation standard does not accurately predict the performance on a 'typical image' especially when image statistics differ greatly from the images in the standard. Although we would like to expand the standard to include more original noise-free images, different sized images, and more noise distributions, we feel that this would increase the size of the noisy image data set beyond a reasonable amount.

The proposed standard currently only contains gray-scale images. Although many denoising algorithms for color images also have an associated gray-scale algorithm and there are far fewer denoising algorithms for color images, there still should be a similar standard for denoising algorithms for color images. In the future, we may develop a similar standard for color images.

# References

[1] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, Image Denoising using a Scale Mixture of Gaussians in the Wavelet Domain, IEEE Transactions on Image Processing, vol. 12, no. 11, pp. 13381351, Nov. 2003.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising with Block-Matching and 3D Filtering," Proc. SPIE Electronic Imaging: Algorithms and Systems V, no. 6064A-30, Jan. 2006.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising by Sparse 3D Transform-Domain Collaborative Filtering," IEEE Transactions on Image Processing, Dec. 2006. (accepted)

[4] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color

Images," IEEE Transactions on Image Processing, vol. 16, no. 5, pp. 1395-1411, May 2007.

[5] M. Elad and M. Aharon, "Image Denoising via Learned Dictionaries and Sparse Representation," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 895-900, Jun. 2006.

[6] M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant representations over Learned Dictionaries," IEEE Transactions on Image Processing, vol. 15, no. 12, pp. 3736-3745, Dec. 2006.

[7] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[8] R. M. Rangayyan, M. Ciuc, and F. Faghih, "Adaptive-Neighborhood Filtering of Images Corrupted by Signal-Dependent Noise," Applied Optics, vol. 37, no. 20, pp. 4477-4487, Jul. 1998.

[9] Z. Long and N. H. Younan, "Denoising of Images with Multiplicative Noise Corruption," Proc. European Signal Processing Conference, Sept. 2005.

[10] R. M. Rangayyan and A. Das, "Filtering Multiplicative Noise in Images using Adaptive Region-Based Statistics," Journal of Electronic Imaging, vol. 7, no. 1, pp. 222-230, Jan. 1998.

[11] K. Hirakawa and T. W. Parks, "Image Denoising for Signal-Dependent Noise," Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, vol. 2, pp. 29-32, Mar. 2005.

[12] Y. Hawwar and A. Reza, "Spatially Adaptive Multiplicative Noise Image Denoising Technique," IEEE Transactions on Image Processing, vol. 11, no. 12, pp. 1397-1404, Dec. 2002.

[13] M. Kovaci, D. Isar, A. Isar, "Denoising SAR Images," International Symposium on Signals, Circuits and Systems, vol. 1, pp. 281-284, Jul. 2003.

[14] X. Huang, A. C. Madoc, and A. D. Cheetham, "Wavelet-Based Bayesian Estimator for Poisson Noise Removal from Images," Proc. International Conference on Multimedia and Expo, vol. 1, pp. 593-596, Jul. 2003.

[15] E. Vansteenkiste, D. Van der Weken, W. Philips, and E. Kerre, "Perceived Image Quality Measurement of State-of-the-Art Noise Reduction Schemes," in Lecture Notes in Computer Science ACIVS, vol. 4179, pp. 114-126, Sept. 2006.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[17] H. R. Sheikh, A. C. Bovik, G. de Veciana, "An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics," IEEE Transactions on Image Processing, vol. 14, no. 12, Dec. 2005.