SSIM-MAXIMIZED LINEAR ESTIMATOR FOR IMAGE DENOISING

Sijie Xiong

Rutgers, The State University of New Jersey

ABSTRACT

In this project, a new image quality measure – the *structural similarity* (SSIM) index is used to design a linear estimator for signals distorted by a zero-mean AWGN channel with known noise variance. The SSIM index better captures perceptual information than the *mean square error* (MSE) measure. And the linear estimator developed w.r.t. the SSIM index performs much better than the *linear least square error* (LLSE) estimator in denoising without additional computational cost.

1. INTRODUCTION

Image quality assessment (IQA) plays a very important role in the design and optimization of image processing algorithms and systems, including image denoising and restoration, equalizer design, contrast enhancement, image approximation, quantization and coding, etc. The state-of-the-art IQA measure is the *mean square error* (MSE). It has desirable properties like convexity, differentiability and distance preservation under orthogonal transformations, which makes the image processing algorithms easy to implement and solve. However, there have also been controversial and critical opinions regarding its weak performance compared to perceptual evaluations by human eyes [1].

To overcome the drawbacks of MSE, Wang et al. [2, 3] proposed an objective IQA metric called *structural similarity* (SSIM) index, which is inspired by the human visual system (HVS), and demonstrated its effectiveness and superiority over MSE [4]. The SSIM measures the quality of a distorted image from the original one, by comparing, locally, the luminance and contrast differences, and structural similarity between them, and averaging these quantities from local patches to yield a global index. Since the ultimate evaluation of images is by the HVS, potentially, optimizing image processing algorithms with respect to the SSIM index should result in better performance than MSE.

In this paper, we study linear estimators optimized with respect to the SSIM index and MSE, for estimating zero-mean Gaussian sources which are distorted by an additive white Gaussian noise (AWGN) channel with zero mean and known variance. The estimators are further utilized to denoise, in the space domain, a natural image also distorted by the AWGN channel. The differences between these two estimators will be investigated and we will validate the improvement in visual quality by the SSIM-optimized linear estimator.

2. MSE VS. SSIM

We provide a brief discussion of the MSE as a signal fidelity measure. Suppose $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$ and $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$ are two signal vectors (e.g., vectorized images), where N is the number of signal samples (e.g., pixels of the image). The MSE between the signals is

$$\mathrm{MSE}(\mathbf{x}, \mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2. \tag{1}$$

The peak signal-to-noise ratio (PSNR) is a variant of the MSE, which is defined as $\mathrm{PSNR} = 10\log_{10}(L^2/\mathrm{MSE})$, where L is the dynamic range of image pixel intensities. For example, 8 bits/pixel gray-scale images will have $L = 255$. The PSNR comes handy if images with different dynamic ranges are to be compared, otherwise it contains no additional information than the MSE. For the purpose of this project, we will be focusing on the MSE measure.

As discussed in the previous section, MSE possesses many fruitful properties for image analysis, but unfortunately, it is unable to capture perceptual information. This can be visualized in Figure 1, in which several distorted "flower" images with approximately the same MSE (≈ 50) are shown in comparison with the original image. We can see that although the MSEs are the same, the distorted images are quite different according to human perception. Meanwhile, the SSIM index values (the larger the better image quality) are also given, which are apparently much more consistent with perceptual evaluations of the distorted images.

Given the two signal vectors $\mathbf{x}$ and $\mathbf{y}$ which are the original and distorted images, the SSIM index [3] is defined as:

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

where the 3 factors

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}, \tag{3}$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}, \tag{4}$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}, \tag{5}$$

Fig. 1: Comparison of "flower" images with different types of distortions, all with MSE $\approx$ 50. From left to right and top to bottom: original image, Gaussian noise (SSIM $=$ 0.2742), salt-and-pepper noise (SSIM $=$ 0.0810), speckle noise (SSIM $=$ 0.7069), blurred image (SSIM $=$ 0.5354), JPEG compressed image (SSIM $=$ 0.6155).

measure the differences between $\mathbf{x}$ and $\mathbf{y}$ in luminance, contrast and structure, respectively, and $\alpha, \beta, \gamma > 0$ are parameters used to adjust the relative importance of them. The quantities μ_x, μ_y, σ_x^2, σ_y^2 and σ_{xy} are summarized in Table 1, and the constants c_1, c_2, c_3 are aimed for numerical stability when the quantities become very small. Note that SSIM $\in [-1, 1]$ with SSIM$(\mathbf{x}, \mathbf{y}) = 1$ if and only if $\mathbf{x} = \mathbf{y}$, so the larger the SSIM value is, the closer the distorted image is to the original image. In this project, we use a simplified form of the SSIM index with $\alpha, \beta, \gamma = 1$ and $c_3 = c_2/2$:

$$\mathrm{SSIM}(\mathbf{x}, \mathbf{y}) = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}\right)\left(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}\right). \quad (6)$$

In the experiment, non-overlapping image patches from the original and distorted images are vectorized to form pairs of $\mathbf{x}$ and $\mathbf{y}$ and to calculate local SSIM index values. Then the local values are averaged to produce an overall quality score of the distorted image, relative to the original one. The simplified SSIM index will be used to design a linear estimator for zero-mean Gaussian source distorted by AWGN channel in the next section. And its performances in estimating the Gaussian source, as well as denoising a natural image (also distorted by AWGN channel) will be presented and compared to the *linear least square error* (LLSE) estimator (developed under the MSE metric) in Section 4.

3. SSIM-MAXIMIZED LINEAR ESTIMATOR

We assume the entries of $\mathbf{x}$ and $\mathbf{y}$ are i.i.d. random variables which can be characterized by X and Y, respectively. The SSIM index given by (6) can be extended to measure the perceptual difference between realizations of X and Y. Suppose

Quantity	Equation
μ_x	$\frac{1}{N}\sum_{i=1}^{N} x_i$
μ_y	$\frac{1}{N}\sum_{i=1}^{N} y_i$
σ_x^2	$\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2$
σ_y^2	$\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu_y)^2$
σ_{xy}	$\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$

Table 1: Quantities used in calculating the SSIM index.

$\mu_x = \mu_y$, then the first term in (6) equals to 1, and the SSIM index becomes

$$\mathrm{SSIM}(X, Y) = \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}. \quad (7)$$

The following theorem [5] gives the SSIM-maximized linear estimator, according to (7), for a Gaussian source distorted by an AWGN channel with zero mean and known variance.

Theorem 1. *For a Gaussian random variable $X \sim \mathcal{N}(0, \sigma_x^2)$, distorted by an uncorrelated AWGN channel noise $N \sim \mathcal{N}(0, \sigma_n^2)$, the linear estimator $\hat{X} = aY + b$ of X after observing $Y = X + N \sim \mathcal{N}(0, \sigma_y^2)$ (where $\sigma_y^2 = \sigma_x^2 + \sigma_n^2$) which maximizes SSIM$(X, \hat{X})$ according to* (7) *is:*

$$\hat{X}_{ssim} = \frac{-c_2\sigma_y^2 + \sqrt{c_2^2\sigma_y^4 + 4\sigma_x^2\sigma_y^2(c_2\sigma_x^2 + \sigma_x^4)}}{2\sigma_x^2\sigma_y^2} Y. \quad (8)$$

Proof. Since X has zero mean, it is trivial that $b = 0$. Substituting $\hat{X} = aY$ in SSIM$(X, \hat{X})$ yields

$$\mathrm{SSIM}(X, \hat{X}) = \frac{2a\sigma_x^2 + c_2}{\sigma_x^2 + a^2\sigma_y^2 + c_2}, \quad (9)$$

setting $\partial \mathrm{SSIM}(X, \hat{X})/\partial a = 0$ gives

$$(\sigma_x^2\sigma_y^2)a^2 + (c_2\sigma_y^2)a - (c_2\sigma_x^2 + \sigma_x^4) = 0. \quad (10)$$

Since $a > 0$, we now have

$$a = \frac{-c_2\sigma_y^2 + \sqrt{c_2^2\sigma_y^4 + 4\sigma_x^2\sigma_y^2(c_2\sigma_x^2 + \sigma_x^4)}}{2\sigma_x^2\sigma_y^2}. \quad (11)$$

$\square$

For $c_2 = 0$, the SSIM-maximized linear estimator in (8) becomes

$$\hat{X}_{ssim} = a_{ssim} = \frac{\sigma_x}{\sigma_y} Y. \quad (12)$$

As is known, the LLSE estimator in the same setting is

$$\hat{X}_{llse} = a_{llse} Y = \frac{\sigma_x^2}{\sigma_y^2} Y. \quad (13)$$

We can see that $a_{ssim} = \sqrt{a_{llse}}$ for $c_2 = 0$ and the computational cost is the same for both estimators.

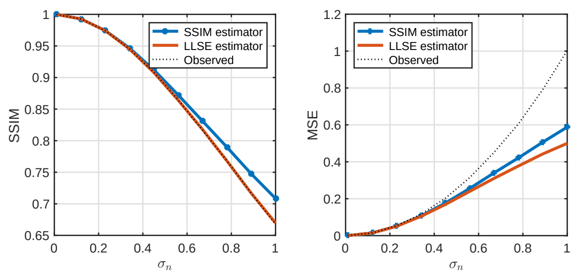


Fig. 2: (Gaussian source denoising) Comparisons of the SSIM-maximized linear estimator and LLSE estimator under the SSIM index and MSE metrics.

4. SIMULATION RESULTS

By applying both linear estimators in the previous section to distorted Gaussian source and natural image, we can compare their performances under SSIM index, MSE and perceptual evaluation.

4.1. Gaussian Source

We generate a random 32×32 signal patch $\mathbf{x}$ (with i.i.d. entries) according to the standard Gaussian distribution $\mathcal{N}(0,1)$, and send it over an AWGN channel $N \sim \mathcal{N}(0, \sigma_n^2)$, where σ_n^2 ranges from 0.01 to 1 and is known to the estimators. The SSIM-maximized linear estimator (12) and the LLSE estimator (13) are then applied to the data points independently from the output patch $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where entries of $\mathbf{n}$ are i.i.d. realizations of N. The quantity σ_y^2 is estimated by the sample variance s_y^2 of the entire patch $\mathbf{y}$, and σ_x^2 is estimated as $s_y^2 - \sigma_n^2$. SSIM$(\mathbf{x}, \hat{\mathbf{x}}_{ssim})$, SSIM$(\mathbf{x}, \hat{\mathbf{x}}_{llse})$, SSIM$(\mathbf{x}, \mathbf{y})$, MSE$(\mathbf{x}, \hat{\mathbf{x}}_{ssim})$, MSE$(\mathbf{x}, \hat{\mathbf{x}}_{llse})$ and MSE$(\mathbf{x}, \mathbf{y})$ are calculated and averaged across 1000 runs and we compare them in Figure 2. When the noise variance is relatively small (less than a half of the signal variance), both linear estimators have the same performance (no improvements over the actual output patch $\mathbf{y}$) under both the SSIM index and MSE metrics. However, if the noise variance becomes large, the SSIM-based linear estimator performs better in terms of SSIM index and the LLSE estimator performs better in terms of MSE, which is trivial. Note that the output patch $\mathbf{y}$ and the LLSE estimator $\hat{\mathbf{x}}_{llse}$ have the same quality score in terms of the SSIM index measure, independent of the noise variance, meaning that although the LLSE estimator reduces the MSE, it doesn't enhance the perceptual image quality at all. This result can also be demonstrated by applying both linear estimators to a natural image distorted by an AWGN channel.

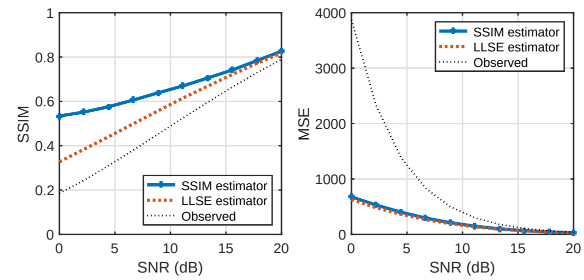


Fig. 3: (Natural image denoising with block size 8×8) Comparisons of the SSIM-maximized linear estimator and LLSE estimator under the SSIM index and MSE metrics.

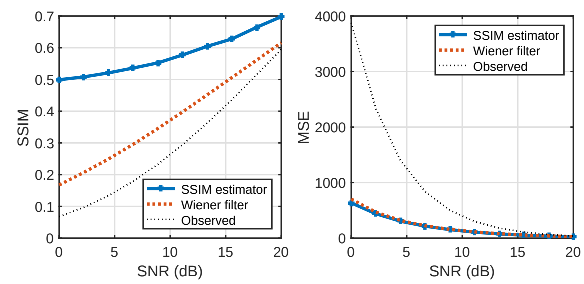
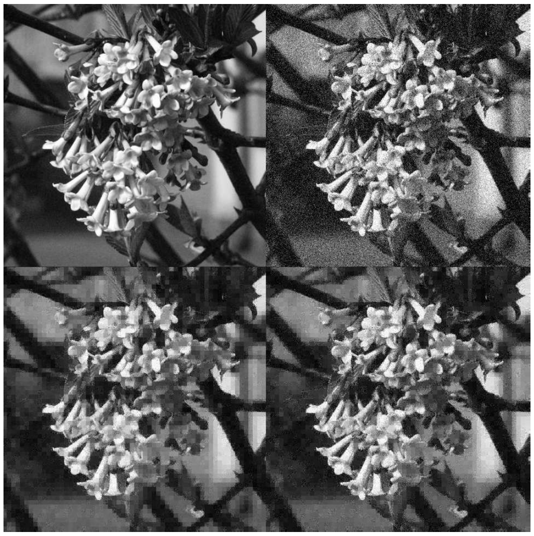


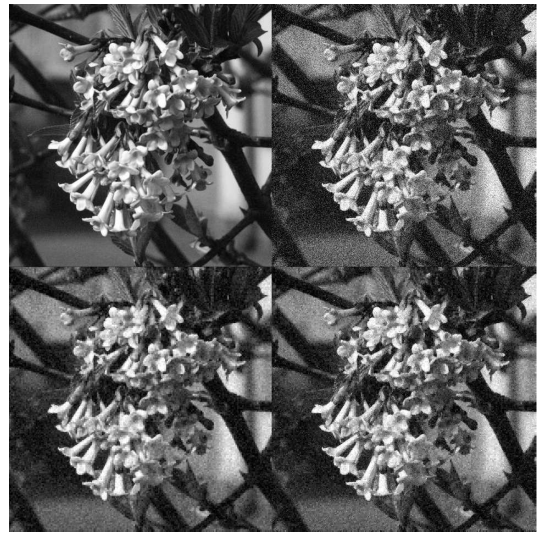
Fig. 4: (Natural image denoising with block size 3×3) Comparisons of the SSIM-maximized linear estimator and LLSE estimator under the SSIM index and MSE metrics.

4.2. Natural Image

To compare the performance of the two estimators in terms of human perception, we use them to denoise, in the space domain, the "flower" image sent over an AWGN channel (with known SNR varying from 0dB to 20dB). Following the same procedure as the Gaussian source denoising, the estimation of a_{ssim}, a_{llse} is performed at a block level. Figure 3 and 4 show the performance comparisons of both linear estimators carried out in different block sizes 8×8 and 3×3, respectively. For block size 3×3, we used the benchmark denoising method – wiener filtering, as the LLSE estimator. We can see that both estimators improve the quality of the noisy image in terms of SSIM index, and the SSIM-maximized linear estimator performs better. Meanwhile, both estimators also enhance the noisy image under the MSE measure, but they have the same performance. These observations can be validated and visualized in Figure 5a and 5b. As we look closely, the SSIM-maximized linear estimator produces better denoised images than the LLSE estimator and wiener filter. Although the LLSE estimator reduces noise, it also introduces larger defects than the SSIM-maximized linear estimator, resembling a lossy compression in some way, which explains the overlapping between SSIM$(\mathbf{x}, \hat{\mathbf{x}}_{llse})$ and SSIM$(\mathbf{x}, \mathbf{y})$ in Figure 2, 3 and 4.



(a) Block size 8×8.



(b) Block size 3×3.

Fig. 5: From left to right and top to bottom: original, noisy, LLSE-denoised and SSIM-denoised images. (SNR=6dB)

5. DISCUSSION

In this project, we studied a new image fidelity measure – the SSIM index, which approximates the human perception better than the canonical MSE measure. By applying a linear estimator, designed to maximize the SSIM index, for a Gaussian source and a natural image distorted by an AWGN channel with zero mean and known noise variance (or SNR), we showed its superiority in performance than the classic LLSE estimator and its consistency with human perceptual evaluations.

We can see in Figure 3 and 4 that different block sizes results in different SSIM index values for both estimators. Specifically, the SSIM index decreases when the block size decreases. How to choose the optimal block size can also be an interesting question to investigate. There have been other image processing tasks (e.g., image restoration [6]) which are designed to optimize the SSIM index instead of MSE. Dominique et. al. [7] also investigated the mathematical properties of the SSIM index which can serve as the foundation for redesigning image optimization problems.

6. REFERENCES

[1] Zhou Wang and Alan C Bovik, "Mean squared error: love it or leave it? a new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[2] Zhou Wang and Alan C Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.

[3] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[4] Zhou Wang, "Applications of objective image quality assessment methods [applications corner]," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 137–142, 2011.

[5] Sumohana S Channappayya, Alan C Bovik, and Robert W Heath, "A linear estimator optimized for the structural similarity index and its application to image denoising," in *2006 International Conference on Image Processing*. IEEE, 2006, pp. 2637–2640.

[6] Sumohana S Channappayya, Alan Conrad Bovik, Constantine Caramanis, and Robert W Heath, "Design of linear equalizers optimized for the structural similarity index," *IEEE transactions on image processing*, vol. 17, no. 6, pp. 857–872, 2008.

[7] Dominique Brunet, Edward R Vrscay, and Zhou Wang, "On the mathematical properties of the structural similarity index," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1488–1499, 2012.