A LINEAR ESTIMATOR OPTIMIZED FOR THE STRUCTURAL SIMILARITY INDEX AND ITS APPLICATION TO IMAGE DENOISING

Sumohana S. Channappayya, Alan C. Bovik, and Robert W. Heath Jr.

The University of Texas at Austin,
Department of Electrical and Computer Engineering,
1 University Station C0803, Austin, TX 78712-0240 USA
{sumohana, bovik, rheath}@ece.utexas.edu

ABSTRACT

We use a perceptual distortion metric - the structural similarity (SSIM) index, to derive a new linear estimator for estimating zero-mean Gaussian sources distorted by additive white Gaussian noise (AWGN). We use this estimator in an image denoising application and compare its performance with the traditional linear least squared error (LLSE) estimator. Although images denoised using the SSIM-optimized estimator have a lower peak signal-to-noise ratio (PSNR) compared to their LLSE couterparts, the SSIM-optimized estimator clearly outperforms the LLSE estimator in terms of the visual quality of the denoised images.

1. INTRODUCTION

For many years, mean squared error (MSE) has been the most popular metric used in the design of image processing algorithms ranging from quantization to image denoising to image quality assessment. This has been the case even though it has been shown that MSE optimized algorithms do not necessarily produce images with the highest visual quality. The popularity of MSE can be attributed to two main reasons : amenability to analysis and a lack of competitive perceptual distortion metrics. Recent advances in full-reference image quality assessment have resulted in the emergence of two powerful perceptual distortion metrics called the SSIM Index [1], or Wang-Bovik Index, and the Visual Information Fidelity Criterion (VIF) [2], or Sheikh-Bovik Index.

The SSIM index computes the quality of a distorted image by comparing the correlations in luminance, contrast, and structure, locally, between the reference and distorted images and averaging these quantities over the entire image. VIF measures image quality as the mutual information between the reference image and the distorted image. The design of both these metrics is inspired by the functioning of the human visual system (HVS). Since most images are intended for viewing by the human eye, using these metrics in designing image processing algorithms automatically takes into account the properties of the HVS. Ideally, optimizing algorithms with respect to the SSIM index and VIF should guarantee the desired level of visual quality. The optimization process is non-trivial however, since the form of these metrics is more complicated than that of MSE.

Several important problems such as estimation for image denoising, rate-distortion analysis, and joint source-channel coding of images can now be solved with respect to these new metrics. In this paper, we propose the first-ever linear estimator optimized with respect to the SSIM index. The choice of the SSIM index as the distortion metric is mainly due to its strength as a perceptual distortion metric, intuitiveness, amenability to analysis, and ease of implementation. We derive an optimal linear estimator with respect to the SSIM index for the simple but important case of estimating zero-mean Gaussian sources that have been distorted by an AWGN channel. The SSIM index defined for the space domain is used in this analysis. The SSIM-optimized estimator is then used to denoise natural images, also distorted by AWGN. The denoiser has been implemented in the space domain in order to be consistent with the flavor of the SSIM index used in the estimator design. We compare the performance of the SSIM-optimal denoiser with the ubiquitous LLSE denoiser and demonstrate the gain in visual quality offered by the proposed method.

2. THE STRUCTURAL SIMILARITY INDEX

We briefly discuss the SSIM visual quality metric to help set up the estimation problem. The most general form of the metric that is used to measure the structural similarity between two signal vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$SSIM(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha [c(\mathbf{x}, \mathbf{y})]^\beta [s(\mathbf{x}, \mathbf{y})]^\gamma. \quad (1)$$

The term $l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$ compares the luminance of the signals, $c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$ compares the contrast of the signals, and $s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$ measures the structural correlation of the signals. The quantities μ_x, μ_y are the sample means of $\mathbf{x}$ and $\mathbf{y}$ respectively, σ_x^2, σ_y^2 are the sample vari-

ances of $\mathbf{x}$ and $\mathbf{y}$ respectively, and σ_{xy} is the sample cross-covariance between $\mathbf{x}$ and $\mathbf{y}$. The constants C_1, C_2, C_3 are used to stabilize the metric for the case where the means and variances become very small. The parameters $\alpha > 0$, $\beta > 0$, and $\gamma > 0$, are used to adjust the relative importance of the three components. We use the following simplified form of the SSIM index in our work (with $\alpha = \beta = \gamma = 1$, and $C_3 = C_2/2$):

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

In image quality assessment, non-overlapping image blocks from the reference and distorted image constitute $\mathbf{x}$ and $\mathbf{y}$ respectively. The average of the SSIM values across the image (also called mean SSIM or MSSIM) gives the final quality measure. The design philosophy of the SSIM index is to acknowledge the fact that natural images are highly structured, and that the measure of structural correlation (between the reference and the distorted image) is very important in deciding the overall visual quality. Further, the SSIM index measures quality locally and is able to capture local dissimilarities better, unlike global quality measures such as MSE (and hence PSNR). Though (2) has a form that is more complicated than that of MSE, it remains analytically tractable. These features make the SSIM index attractive to work with.

3. SSIM-OPTIMIZED LINEAR ESTIMATION

The first step in designing SSIM-optimized estimators is to extend the defintion of the SSIM index to measure structural similarity between realizations of random variables. To do so, the same formulation as in (2) is used, but sample means, variances, and cross-covariance are replaced by their statistical equivalents. For zero-mean reference and distorted signals, the luminance term in (2) goes to unity and the structure term remains and the SSIM index becomes

$$SSIM(X, Y) = \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}. \quad (3)$$

Based on (3) we can prove the following theorem.

Theorem 3.1. *For a zero-mean Gaussian random variable $X \sim \mathcal{N}(0, \sigma_x^2)$ that is sent over an uncorrelated AWGN channel $N \sim \mathcal{N}(0, \sigma_n^2)$, the linear estimator $\hat{X} = f(Y) = aY + b$ that maximizes (3) is given by*

$$\hat{X}_{ssim} = \frac{-C_2\sigma_y^2 + \sqrt{C_2^2\sigma_y^4 + 4\sigma_x^2\sigma_y^2(\sigma_x^4 + C_2\sigma_x^2)}}{2\sigma_x^2\sigma_y^2}Y, \quad (4)$$

where $Y = X + N$, $Y \sim \mathcal{N}(0, (\sigma_y^2 = \sigma_x^2 + \sigma_n^2))$, is the observed random variable at the receiver.

Proof : It is easy to show that for zero-mean random variables, $b = 0$. Substituting $\hat{X}$ with aY in (3) and maximizing the expression with respect to a yields the optimal linear
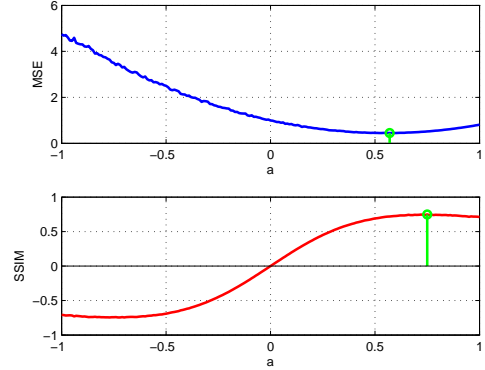
Fig. 1. Plot of end-to-end MSE and SSIM as a function of the linear estimator a for a zero-mean Gaussian source with $\sigma_x = 1$ sent over an AWGN channel with $\sigma_n = 0.9$. The plot also shows the minimum MSE and maximum SSIM points.

SSIM estimator in (4). If the constant $C_2 = 0$, (4) simplifies to

$$\hat{X}_{ssim} = a_{ssim}Y = \frac{\sigma_x}{\sigma_y}Y. \quad (5)$$

$\square$

The estimator in (4) is also the optimal linear estimator for the case where the means of X and Y are identical.

For Gaussian sources, the LLSE estimator is also the minimum MSE (MMSE) estimator. For the case of zero-mean Gaussian random variables, the LLSE estimator is

$$\hat{X}_{llse} = a_{llse}Y = \frac{\sigma_x^2}{\sigma_y^2}Y. \quad (6)$$

The expressions for MSE and the SSIM index for the SSIM-optimal and LLSE estimators in (5) and (6) respectively are

$$\begin{aligned}
MSE_{ssim} &= 2\sigma_x^2(1 - a_{ssim}), \\
MSE_{llse} &= \sigma_x^2(1 - a_{llse}), \\
SSIM_{ssim} &= 2a_{ssim}, \\
SSIM_{llse} &= [2a_{llse}/(1 + a_{llse})].
\end{aligned} \quad (7)$$

To elucidate the effect of the linear estimator (aY) on the SSIM index and MSE, we ran the following experiment. The signal and noise variance σ_x^2 and σ_n^2 were chosen to be 1 and 0.81 respectively (this choice was made to highlight the differences between the estimators). The value of a was varied from -1.0 to 1.0 in steps of 0.01 and the estimator $\hat{X} = aY$ was applied at the receiver. The end-to-end SSIM index and MSE were measured as shown in Fig. 1. The figure also shows the points where MSE is minimum and the SSIM index is maximum. It is clear that for this SNR, the values of a that minimize MSE and maximize the SSIM index are quite different. Eqns. (4), (5), and (6) suggest that for zero-mean Gaussian sources, as $C_2 \to 0$, $a_{ssim} \approx \sqrt{a_{llse}}$.

(a) SSIM vs. Noise variance

(b) MSE vs. Noise variance
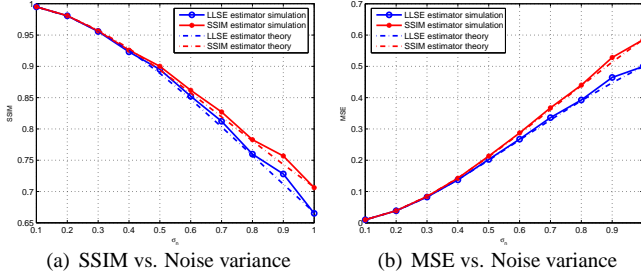
Fig. 2. SSIM and MSE plots for the optimal linear SSIM estimator and LLSE estimator for a zero-mean unit variance Gaussian patch of size 32×32 that is sent over an AWGN channel whose variance is varied from 0.01 to 1.

(a) SSIM vs. Channel SNR

(b) PSNR vs. Channel SNR

Fig. 3. SSIM and PSNR plots for the optimal linear SSIM estimator and LLSE estimator applied at the block-level to Img0099 from the 'Images of Austin' database.

4. PERFORMANCE COMPARISON
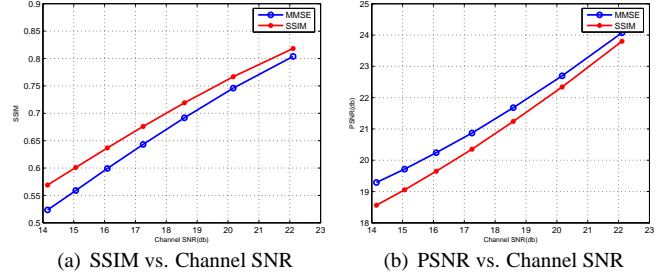
4.1. Gaussian Source

A zero-mean, unit variance Gaussian patch of size 32×32 was sent over an AWGN channel N, whose variance was varied from 0.01 to 1 in steps of 0.1. The SSIM-optimized and LLSE estimators were independently applied to the received patch Y, and the end-to-end SSIM was measured for each case. Each point in the plots in Fig. 2 has been averaged over 10,000 iterations. We see from Fig. 2(a) that the SSIM-optimized estimator starts to perform better (with respect to SSIM) as the signal-to-noise ratio decreases. This seems to suggest that the SSIM estimator should give perceptually superior images compared to the LLSE estimator as the noise variance increases. Fig. 2(b) shows how the end-to-end MSE varies as the noise variance increases. As expected, the LLSE estimator outperforms the SSIM-optimized linear estimator with respect to MSE.

4.2. Natural Images

We compared the performance of the two estimators in spatial domain denoising of natural images distorted by an AWGN channel. Estimation was performed at a *block level* (since the SSIM index performs block level comparison), and at a *pixel level* (to enable performance comparison with the Wiener filter). Spatial domain denoising was done since the flavor of SSIM used in Section 3 is defined for the same. We assumed that the noise variance was known to the receiver at the block level and the pixel level for the respective cases.

Block-level denoising : In this case the image was divided into non-overlapping blocks of size 8×8 and the SSIM-optimal and LLSE estimators were calculated for each block. The source variance was approximated from the distorted image variance and the noise variance as $\sigma_x^2 \approx \sigma_y^2 - \sigma_n^2$. The estimators were calculated using expressions (4) and (6). Since the optimal estimators were derived for the zero-mean case,

we subtracted the mean of each block before applying the estimator and then added the mean back. The SSIM-optimal estimator consistently outperforms the LLSE estimator in terms of visual quality from block size 4×4 onward. The choice of using an 8×8 block was made to highlight the gain in visual quality offered by the SSIM-optimal estimator.

Pixel-level denoising : We performed pixel level denoising to compare the performance of the SSIM-optimal estimator to the benchmark Wiener filter. The main difference from the block-level denoising case was that we used the function *wiener2* available in $\mathrm{Matlab}^{\circledR}$ to do the LLSE estimation. We used the same neighborhood size to calculate σ_y^2 for both estimators.

The standard set of test images including Lena, Barbara, Peppers, and the 'Images of Austin' database consisting of 200 grayscale, 512×512, bitmap images, of the city of Austin were used in the experiment. The AWGN channel's SNR was varied from 14 dB ($\sigma = 50$) to 22 dB ($\sigma = 20$). At each SNR, the received image was denoised using both estimators, and the PSNR and SSIM of the denoised images was calculated. The PSNR and the SSIM index values at each SNR were averaged over 10 iterations. Fig. 3 shows the plot of PSNR and the SSIM index versus channel SNR for block-level denoising for the sample image Img0099.

5. RESULTS

We see from Fig. 3 that the estimators behave as expected for natural images too. The SSIM- optimized estimator gives consistently higher end-to-end SSIM index values compared to the LLSE estimator, and vice-versa. Most importantly, we see from Figs. 4(a) and 4(b), that the SSIM-optimized estimator gives images that have a higher visual quality. Further, this result also demonstrates that higher PSNR values do not always mean images with a higher perceptual quality. Comparing (5) and (6) we see that the cost incurred (in terms of the number of multiplication operations) in computing the SSIM-optimized estimator is identical to the cost of computing the

(a) Block-level denoising (block size 8×8).
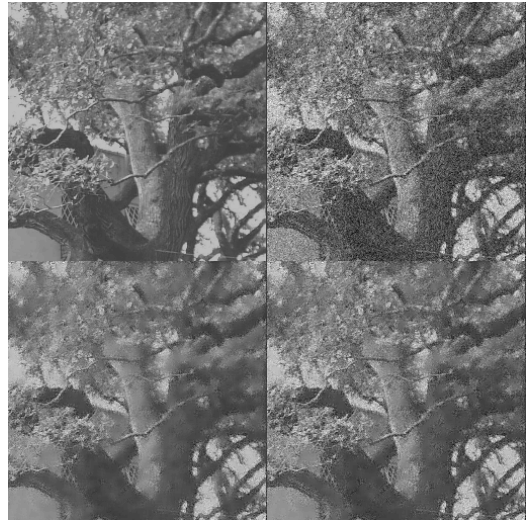
(b) Pixel-level denoising.

Fig. 4. From left to right and top to bottom : Original, Noisy, LLSE, SSIM (cropped to 128×128 for easier comparison) . Left : Barbara image. Noisy image PSNR = 20.2 dB ($\sigma = 25$). $\text{PSNR}_{\text{llse}}$ = 25.13 dB, $\text{PSNR}_{\text{ssim}}$ = 24.66 dB. $\text{SSIM}_{\text{llse}}$ = 0.70, $\text{SSIM}_{\text{ssim}}$ = 0.72. Right : Img0099. Noisy image PSNR = 17.2 dB ($\sigma = 35$). $\text{PSNR}_{\text{wiener}}$ = 21.22 dB, $\text{PSNR}_{\text{ssim}}$ = 20.32 dB. $\text{SSIM}_{\text{wiener}}$ = 0.64, $\text{SSIM}_{\text{ssim}}$ = 0.67.

LLSE estimator. In other words, the gain in visual quality is available at no additional computational cost!

To validate the visual quality results, a subjective study was conducted [3]. The experiment for natural images yields a pair of denoised images at each SNR, for every image in the database, forming a set of 1400 pairs. For brevity, a subset of 10 image pairs was randomly chosen from this set. These image pairs were independently shown to six subjects, and were asked to choose the image with the higher visual quality. On average, 85% of the images chosen by the subjects were those denoised by the SSIM estimator. The worst case of 50% occured only for one image, while the best case of 100% occured for 5 images.

6. CONCLUSIONS

We proposed the first-ever SSIM-optimized linear estimator and demonstrated the improvement in visual quality that results from using the SSIM index in designing estimators for denoising. This paper only considered the special case of a zero-mean Gaussian source and an AWGN channel, but we plan to extend this work to more sophisticated source and channel models. Similar improvements can be expected in state-of-the-art denoisers [4] by optimizing them for the SSIM index instead of a least squared error metric. A flavor of the SSIM index defined in the wavelet domain called the complex wavelet SSIM (CWSSIM) [5] index can be used in wavelet

domain denoising. Furthermore, we believe that the use of this metric is not limited to denoising applications, and can be used in other important applications such as quantizer design, and the design of joint source-channel coding systems.

7. REFERENCES

[1] 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, Apr. 2004.

[2] Hamid R. Sheikh, Alan C. Bovik, and Gustavo de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, pp. 2117–2128, Dec. 2005.

[3] Sumohana S. Channappayya, "Subjective evaluation of images denoised using SSIM-optimal linear estimators and LLSE estimators," http://www.ece.utexas.edu/˜sumohana/ssim.

[4] Javier Portilla, Vasily Strela, Martin J. Wainwright, and Eero P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[5] Zhou Wang and Eero P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005, vol. 2, pp. 573–576.