

SSIM-OPTIMAL LINEAR IMAGE RESTORATION

Sumohana S. Channappayya, Alan C. Bovik, Constantine Caramanis, 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

ABSTRACT

In this paper, we present an algorithm for designing a linear equalizer that is optimal with respect to the structural similarity (SSIM) index. The optimization problem is shown to be non-convex, thereby making it non-trivial. The non-convex problem is first converted to a quasi-convex problem and then solved using a combination of first order necessary conditions and bisection search. To demonstrate the usefulness of this solution, it is applied to image denoising and image restoration examples. We show using these examples that optimizing equalizers for the SSIM index does indeed result in higher perceptual image quality compared to equalizers optimized for the ubiquitous mean squared error (MSE).

Index Terms— Image restoration.

1. INTRODUCTION

The mean squared error (MSE) has been a popular metric used not only to assess the quality of images, but also in optimizing various image processing algorithms. It has been shown however, that the MSE is not the best metric either for quality assessment or for optimizing image processing algorithms [1]. The MSE is popular because it lends itself well to analysis, and due to a lack of competitive image quality assessment (IQA) algorithms. Recent advances in full-reference IQA has resulted in a number of powerful new algorithms such as the SSIM index [2], the visual information fidelity (VIF) criterion [3], and the visual signal to noise ratio [4]. In this paper, we restrict our focus to the SSIM index.

The SSIM index has been shown to outperform MSE and the related PSNR in measuring the quality of natural images across a wide variety of distortions [2]. 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.

Based on the performance of the SSIM index as a powerful IQA algorithm, using it as the objective function in optimizing image processing algorithms appears very promising. This optimization is not straightforward however, given the form of the SSIM index, and algorithms that explicitly optimize for it are only recently being developed. The design of a linear estimator optimized for the SSIM index and its application to image denoising demonstrated the gain in perceptual quality of the denoised images compared to traditional MSE optimal linear estimation [5]. An iterative SSIM optimal bit allocation technique for image coding has also shown promising gains in the perceptual quality of the compressed images [6].

In this paper, we formulate and solve the general problem of designing a linear equalizer of length L that is optimized with respect to the SSIM index. The definition of the SSIM index is extended to wide sense stationary (WSS) sources, and the optimiza-

tion is carried out using this extended definition. The main contribution of this paper is the solution to the non-convex optimization problem. The equalizer is then applied to both image denoising and image restoration problems. The results demonstrate that using SSIM-optimal equalizers for image denoising and restoration results in higher perceptual quality of the restored images when compared to traditional MSE-optimal solutions.

We begin by a brief introduction to the SSIM index and extend its definition to WSS processes. The equalization problem is then formulated and shown to be non-convex. The problem is converted to a quasi-convex problem and its solution is presented. Finally, the results are applied to image denoising and restoration problems.

2. THE SSIM INDEX

The most general form of the metric that is used to measure the structural similarity between two signal vectors \mathbf{x} and \mathbf{y} in \mathbf{R}^n 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 mean 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 variance of the signals, and $s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$ measures the 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 variances 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, pixel values of local image patches from the reference and distorted image constitute \mathbf{x} and \mathbf{y} respectively. The term $l(\mathbf{x}, \mathbf{y})$ compares the luminance, $c(\mathbf{x}, \mathbf{y})$ compares the contrast, and $s(\mathbf{x}, \mathbf{y})$ compares the structure of the local image patches. The average of the SSIM values across the image (also called mean SSIM or MSSIM) gives the final quality measure. The key idea behind 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 MSE, it remains analytically tractable.

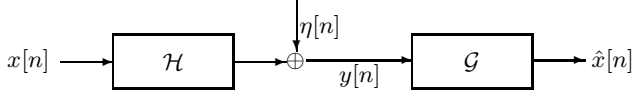


Fig. 1. Block diagram of a general linear time invariant system. The design objective is to find the linear equalizer block \mathcal{G} , given the observed process $y[n]$, the LTI filter \mathcal{H} , and the power spectral density of the noise process $\eta[n]$ so that the StatSSIM index between $x[n]$ and $\hat{x}[n]$ is maximized.

3. PROBLEM FORMULATION

In this section, we first extend the definition of the SSIM index to measure similarity between WSS random processes. The equalization problem is formulated using the extended definition of the SSIM index. It is shown that the problem is a non-convex function of the equalizer coefficients.

Definition Given two WSS random processes $x[n]$ and $y[n]$ with means μ_x and μ_y respectively, the statistical SSIM index is defined as

$$\text{StatSSIM}(x[n], y[n]) = \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right) \left(\frac{2E[(x[n] - \mu_x)(y[n] - \mu_y)] + C_2}{E[(x[n] - \mu_x)^2] + E[(y[n] - \mu_y)^2] + C_2} \right). \quad (3)$$

This is a straightforward extension of the pixel domain definition of the SSIM index by replacing sample means and variances with their statistical equivalents.

3.1. Equalization Problem

The equalization problem is illustrated in Fig. 1. We assume that the input to the system $x[n]$ is a WSS process, $h[n]$ is a linear time invariant (LTI) filter known at the receiver, and the noise process $\eta[n]$ is white and its power spectral density (PSD) is known at the receiver. The following problem is considered. Given a distorted observation $y[n] = h[n] * x[n] + \eta[n]$ of the input process $x[n]$, design a filter $g[n]$ of length N such that the StatSSIM index between the reference $x[n]$ and the restored process $\hat{x}[n] = g[n] * y[n]$ is maximized (where $*$ is the convolution operator).

Rewriting the StatSSIM index as a function of $g[n]$ using (3) we get

$$\text{StatSSIM}(x[n], \hat{x}[n]) = \left(\frac{2\mu_x \mathbf{g}^T \mathbf{e} \mu_y + C_1}{\mu_x^2 + \mathbf{g}^T \mathbf{e} \mathbf{e}^T \mathbf{g} \mu_y^2 + C_1} \right) \left(\frac{2\mathbf{g}^T \mathbf{c}_{xy} + C_2}{\sigma_x^2 + \mathbf{g}^T \mathbf{K}_{yy} \mathbf{g} + C_2} \right) \quad (4)$$

where $\mathbf{g} = [g[0], g[1], \dots, g[N-1]]^T$, $\mathbf{e} = [1, 1, \dots, 1]^T$ are both length N vectors, μ_x, μ_y are the means of the source and observed processes respectively, $\mathbf{c}_{xy} = E[(x[n] - \mu_x)(y[n] - \mu_y)]$, is the cross-covariance between the source ($x[n]$) and the observed processes ($\mathbf{y} = [y[n], y[n-1], \dots, y[n-(N-1)]]^T$), σ_x^2 is the variance of the source process at zero delay, $\mathbf{K}_{yy} = E[(\mathbf{y} - \mathbf{e}\mu_y)(\mathbf{y} - \mathbf{e}\mu_y)^T]$, is the covariance matrix of size $N \times N$ of the observed process $y[n]$, and C_1, C_2 are stabilizing constants. The equalization problem is expressed as

$$\mathbf{g}^* = \operatorname{argmax}_{\mathbf{g} \in \mathbb{R}^N} \text{StatSSIM}(x[n], \hat{x}[n]). \quad (5)$$

4. STATSSIM-OPTIMAL LINEAR EQUALIZATION

From (4) we see that the StatSSIM index is a non-convex function of the equalizer coefficients \mathbf{g} (the numerator is a second degree polynomial in \mathbf{g} , while the denominator is a fourth degree polynomial in \mathbf{g}). This implies that local optimality conditions such as Karush-Kuhn-Tucker (KKT) cannot guarantee global optimality. In particular, any approach based on descent-type algorithms are likely to get stuck in local optima. The approach that we take to solve this problem transforms the non-convex problem into a quasi-convex formulation. Convex optimization problems are efficiently solvable using widely available optimization techniques and software [7, 8]. We show, moreover, that in addition to the convex reformulation, we can obtain a near-closed form solution. In particular, we reduce the N -tap filter optimization, for *any* N , into an optimization problem over only *two* variables. Exploiting convexity properties, we can quickly search over one parameter by means of a bisection technique, thus reducing the problem to a univariate optimization problem. This last step can be quickly performed by means of an analytic solution of a simplified problem, which brings us close to the optimal value of the final variable of our optimization.

4.1. Problem Reformulation

Note that the first term of (4) (corresponding to the mean) is a function only of the sum of the filter coefficients $\mathbf{g}^T \mathbf{e}$. We use this property to simplify the optimization problem in (4) by constraining $\mathbf{g}^T \mathbf{e} = \alpha$. With this constraint, the optimization problem simplifies to finding

$$\left[\begin{array}{l} \mathbf{g}(\alpha) = \operatorname{argmax}_{\mathbf{g} \in \mathbb{R}^N} \left(\frac{2\mathbf{g}^T \mathbf{c}_{xy} + C_2}{\sigma_x^2 + \mathbf{g}^T \mathbf{K}_{yy} \mathbf{g} + C_2} \right), \\ \text{s.t. : } \mathbf{g}^T \mathbf{e} = \alpha. \end{array} \right] \quad (6)$$

This problem is a function of α . The overall problem is to find the highest StatSSIM index by searching over a range of α (typically in the interval $[1 - \delta, 1 + \delta]$, for a small δ). The solution to the optimization problem in (6) is presented in the following section, along with an efficient search strategy for finding α .

4.2. Solution

The maximization problem in (6) is still non-convex. We convert it into a quasi-convex optimization problem as follows

$$\begin{aligned} & \mathbf{g}(\alpha) = \operatorname{argmax}_{\mathbf{g} \in \mathbb{R}^N} \left(\frac{2\mathbf{g}^T \mathbf{c}_{xy} + C_2}{\sigma_x^2 + \mathbf{g}^T \mathbf{K}_{yy} \mathbf{g} + C_2} \right), \\ & \text{s.t. : } \mathbf{g}^T \mathbf{e} = \alpha, \\ & \Leftrightarrow \\ & \min : \gamma \\ & \text{s.t. : } \left[\begin{array}{l} \max : \left(\frac{2\mathbf{g}^T \mathbf{c}_{xy} + C_2}{\sigma_x^2 + \mathbf{g}^T \mathbf{K}_{yy} \mathbf{g} + C_2} \right) \leq \gamma \\ \text{s.t. : } \mathbf{g}^T \mathbf{e} = \alpha, \end{array} \right] \quad (7) \\ & \Leftrightarrow \\ & \min : \gamma \\ & \text{s.t. : } \left[\begin{array}{l} \min : [\gamma(\sigma_x^2 + \mathbf{g}^T \mathbf{K}_{yy} \mathbf{g} + C_2) - (2\mathbf{g}^T \mathbf{c}_{xy} + C_2)] \geq 0 \\ \text{s.t. : } \mathbf{g}^T \mathbf{e} = \alpha \end{array} \right]. \end{aligned}$$

The first step in the reformulation is the introduction of the auxiliary variable γ as an upper bound on (6). The first equivalence relation holds since minimizing γ is the same as finding the least upper bound of the function in (6). This is equal to the maximum value

of the function, which exists, as seen by straightforward continuity arguments. The second equivalence relation holds since the denominator in (6) is strictly positive, allowing us to multiply through and rearrange terms. Then, γ is a true upper bound if the problem

$$\left[\begin{array}{l} \max_{\mathbf{g} \in \mathbb{R}^N} : \quad \gamma(\sigma_x^2 + \mathbf{g}^T \mathbf{K}_{yy} \mathbf{g} + C_2) - (2\mathbf{g}^T \mathbf{c}_{xy} + C_2) \\ \text{s.t.} : \quad \mathbf{g}^T \mathbf{e} = \alpha \end{array} \right] \quad (8)$$

has a non-negative optimal value. Since the objective function is a linear term minus a convex quadratic, it is concave. The constraint is affine, and thus convex. Therefore the overall problem is convex, and can be solved by introducing a Lagrange multiplier λ and applying the first order sufficiency condition s

$$\nabla_{\mathbf{g}} \{ \gamma(\sigma_x^2 + \mathbf{g}^T \mathbf{K}_{yy} \mathbf{g} + C_2) - (2\mathbf{g}^T \mathbf{c}_{xy} + C_2) + \lambda(\mathbf{g}^T \mathbf{e} - \alpha) \} = 0. \quad (9)$$

Solving for \mathbf{g} and λ , and denoting them $\mathbf{g}(\alpha), \lambda(\alpha)$ to emphasize their dependence on α , we have

$$\begin{aligned} \mathbf{g}(\alpha) &= \frac{1}{2\gamma} \mathbf{K}_{yy}^{-1} (2\mathbf{c}_{xy} - \lambda(\alpha)\mathbf{e}) \\ \lambda(\alpha) &= \frac{1}{\mathbf{e}^T \mathbf{K}_{yy}^{-1} \mathbf{e}} (2\mathbf{c}_{xy}^T \mathbf{K}_{yy}^{-1} \mathbf{e} - 2\gamma\alpha). \end{aligned} \quad (10)$$

The optimal γ can then be computed in $O(\log(1/\epsilon))$ iterations using a standard bisection procedure. Such an algorithm is summarized in Fig. 2. In this procedure, the threshold ϵ determines the tightness of the bound γ . In other words, the solution obtained using this technique will be within ϵ of the optimal solution. The efficiency of the algorithm can be improved using better search techniques.

4.3. Search for α

The solution in (10) maximizes the function in (7) to give $\mathbf{g}(\alpha)$, i.e., it is still a function of α . The optimal solution to (4) is found by searching over α . The search is over a bounded one-dimensional interval, and is therefore easy to perform. We present two ways to speed up this search.

The first is to simply initialize α to the sum of the filter coefficients of the MSE-optimal filter, i.e., $\alpha_{init} = \mathbf{g}_{mse}^T \mathbf{e}$. The second is a heuristic technique that was found to work better than the first in all our experiments with natural images. In this technique, α is initialized to the sum of the filter coefficients of a structure-optimal filter. By structure-optimal filter, we mean a filter that optimizes only the structure term in the StatSSIM index without any constraints on the mean. This would yield a filter that is optimal with respect to one of the two terms in the StatSSIM index (4). In the following, the structure-optimal filter is derived.

Following the notation in Section 3.1, our goal is to find a filter \mathbf{g}_{struct}^* such that

$$\mathbf{g}_{struct}^* = \operatorname{argmax}_{\mathbf{g} \in \mathbb{R}^N} \left(\frac{2\mathbf{g}^T \mathbf{c}_{xy} + C_2}{\sigma_x^2 + \mathbf{g}^T \mathbf{K}_{yy} \mathbf{g} + C_2} \right). \quad (11)$$

This problem has the same form as (6), and thus can be quickly solved using the optimization technique given above. The optimal solution is $\mathbf{g}_{struct}^* = \frac{1}{\gamma_{struct}^*} (\mathbf{K}_{yy})^{-1} \mathbf{c}_{xy}$, and so the initial value of α is $\alpha_{init} = \mathbf{e}^T \mathbf{g}_{struct}^*$. The value of γ_{struct}^* is computed using the same algorithm as in Section 4.2, and this value is potentially different from the γ in Section 4.2.

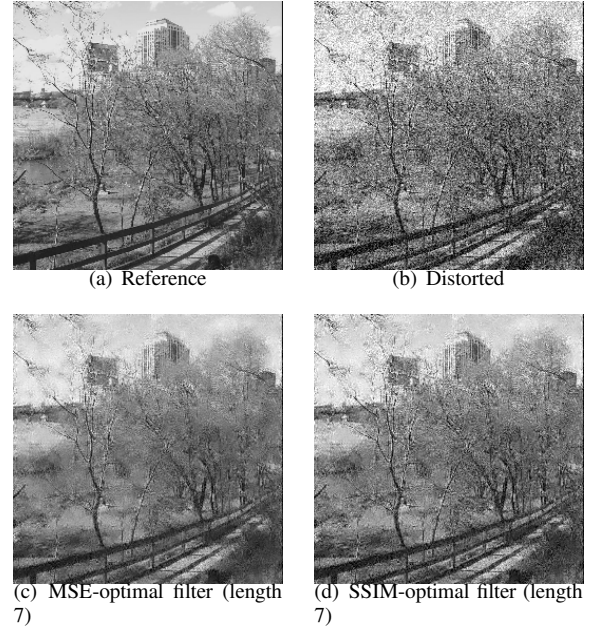


Fig. 3. Img0039.bmp from the ‘City of Austin’ database. 3(a) Original image. 3(b) Distorted image with $\sigma_{noise} = 35$, MSE = 1226.3729, SSIM index = 0.5511. 3(c) Image denoised with a 7-tap MSE-optimal filter, MSE = 436.6929, SSIM index = 0.6225. 3(d) Image denoised with a 7-tap SSIM-optimal filter, MSE = 528.0777, SSIM index = 0.6444.

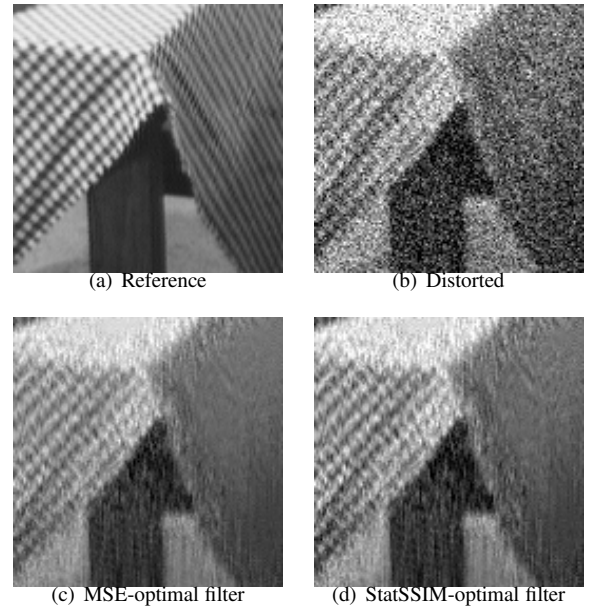


Fig. 4. A 128×128 block of Barbara image. 4(a) Original image. 4(b) Distorted image with $\sigma_{blur} = 1, \sigma_{noise} = 40$, MSE = 1781.9058, SSIM index = 0.5044. 4(c) Image restored with a 11-tap MSE-optimal filter, MSE = 520.1322, SSIM index = 0.6302. 4(d) Image restored with a 11-tap SSIM-optimal filter, MSE = 584.9232, SSIM index = 0.6568.

1. Pick an initial guess of γ (say γ_0) between 0 and 1. `upper_limit = 1, lower_limit = γ_0 .`
2. Evaluate the optimal filter.
3. Is $\gamma(\sigma_x^2 + \mathbf{g}^T \mathbf{K}_{yy} \mathbf{g} + C_2) - (2\mathbf{g}^T \mathbf{c}_{xy} + C_2) \geq 0$?
 - 3a. If true, is `(upper_limit - lower_limit < ϵ)`?
 - 3aa. If true, we have found a γ within ϵ of the optimal value. `Exit.`
 - 3ab. If false, set $\gamma_i = 0.5 * (\text{upper_limit} + \text{lower_limit})$, `upper_limit = γ_i . Goto step 2.`
 - 3b. If false, set $\gamma_i = 0.5 * (\text{upper_limit} + \text{lower_limit})$, `lower_limit = γ_i . Goto step 2.`

Fig. 2. An algorithm to search for the optimal γ .

5. RESULTS

We now apply the SSIM-optimal linear equalizer solution to both image denoising and image restoration examples. The solution in the previous section assumes 1-D signals. For consistency, the solution is applied to vectorized local image patches. Further, the solution is implemented in the image domain. The following subsections provides implementation details for each example.

5.1. Denoising results

For the denoising case, the required statistics are computed as follows. The cross-covariance values at each pixel location are computed from the $N^2 \times 1$ vector formed from a local neighborhood around the given pixel of the observed image. The following relations are used in the computation.

$$c_{xx}(0) = c_{yy}(0) - \sigma_\eta^2. \quad (12)$$

The first equality holds since the noise is zero-mean, therefore, $\mu_y = \mu_x$, and the noise is independent of the source. Similarly, for $\tau \neq 0$,

$$\begin{aligned} c_{yy}(\tau) &= E[(y[n] - \mu_y)(y[n - \tau] - \mu_y)] = c_{xx}(\tau), \\ \Rightarrow c_{xy}(\tau) &= c_{xx}(\tau) = c_{yy}(\tau). \end{aligned} \quad (13)$$

The first equality relation holds since $\eta[n]$ and $\eta[n - \tau]$ are independent (of each other, and the source process $x[n]$) and zero-mean. The second equality relation easily follows from the first.

The results of image denoising are shown in Fig. 3. SSIM-optimal denoising is compared to MSE-optimal denoising, and we see that Fig. 3(d) has a higher perceptual image quality when compared to Fig. 3(c).

5.2. Restoration results

For the restoration case, we follow the heuristic technique presented in [9] to calculate the required image statistics. For brevity, we omit the details of this calculation and refer the reader to [9]. As with the denoising solution, we see that the SSIM-optimal equalizer results in a higher perceptual image quality of the restored image in Fig. 4(d) when compared to the MSE-optimally restored image in Fig. 4(c).

6. CONCLUSIONS

In this paper, we formulated the problem of SSIM-optimal linear equalization. The non-convex problem was re-cast into a quasi-convex form. This allowed for better tractability and admitted a

unique solution. The proposed solution is near closed-form and its complexity is independent of the length of the equalizer. The SSIM-optimal equalizer was then applied to image denoising and restoration examples. Through these examples we showed that optimizing algorithms for the SSIM index does indeed result in improved visual quality of the denoised and restored images compared to their MSE optimally denoised and restored counterparts.

7. REFERENCES

- [1] Bernd Girod, "What's Wrong with Mean-squared Error?," in *Digital Images and Human Vision*, Andrew B. Watson, Ed., pp. 207–220. Cambridge, MA: MIT Press, 1993.
- [2] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [3] Hamid R. Sheikh and Alan C. Bovik, "Image Information and Visual Quality," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [4] Damon M. Chandler and Sheila S. Hemami, "VSNR: A Wavelet-based Visual Signal-to-Noise Ratio for Natural Images," *IEEE Trans. Image Processing*, vol. 16, pp. 2284–2298, 2007.
- [5] Sumohana S. Channappayya, Alan C. Bovik, and Robert W. Heath Jr., "A Linear Estimator Optimized for the Structural Similarity Index and its Application to Image Denoising," in *Proc. of IEEE Intl. Conference on Image Processing*, 2006, vol. 12.
- [6] Zhou Wang, Q. Li, and X. Shang, "Perceptual Image Coding Based on a Maximum of Minimal Structural Similarity Criterion," in *IEEE International Conference on Image Processing*, Sept. 2007.
- [7] Dmitri P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 1995.
- [8] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
- [9] Javier Portilla and Eero P. Simoncelli, "Image Restoration using Gaussian Scale Mixtures in the Wavelet Domain," in *IEEE International Conference on Image Processing*, 2003, vol. 2, pp. 965–968.