IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP11), Prague, Czech, May 2011

SSIM-INSPIRED IMAGE DENOISING USING SPARSE REPRESENTATIONS

Abdul Rehman[1], Zhou Wang[1], Dominique Brunet[2], Edward R. Vrscay[2]

[1]Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada
[2]Dept. of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada
abdul.rehman@uwaterloo.ca, zhouwang@ieee.org, dbrunet@uwaterloo.ca, ervrscay@uwaterloo.ca

ABSTRACT

Perceptual image quality assessment (IQA) and sparse signal representation have recently emerged as high-impact research topics in the field of image processing. Here we make one of the first attempts to incorporate the structural similarity (SSIM) index, a promising IQA measure, into the framework of optimal sparse signal representation and approximation. In particular, we introduce a novel image denoising scheme where a modified orthogonal matching pursuit algorithm is proposed for finding the best sparse coefficient vector in maximum-SSIM sense for a given set of linearly independent atoms. Furthermore, a gradient descent algorithm is developed to achieve SSIM-optimal compromise in combining the input and sparse dictionary reconstructed images. Our experimental results show that the proposed method achieves better SSIM performance and provide better visual quality than least square optimal denoising methods.

Index Terms— SSIM-based approximation, image denoising, sparse representation, structural similarity index, orthogonal matching pursuit

1. INTRODUCTION

Apart from the popularity and ease of use in optimization problems, mean squared error (MSE) is not the best choice when it comes to image quality assessment (IQA) and signal approximation tasks [1]. Among the recently proposed IQA approaches, the structural similarity (SSIM) index [2] has emerged as a promising measure that shows superior performance as compared to MSE [1]. The SSIM index and its extensions have found a wide variety of applications, ranging from image coding, restoration and fusion, to watermarking and biometrics [1]. In most existing works, however, SSIM has been used for quality evaluation and algorithm comparison purposes only. Much less has been done on using SSIM as an optimization criterion in the design and optimization of image processing algorithms and systems [3], [4].

Image denoising is a classical problem of particular interest to image processing researchers, not only for its practical value, but also because it provides an excellent test bed for image modeling, representation and estimation theories. Recently, a highly effective approach, known as K-SVD, is proposed to tackle this problem, which obtained state-of-the-art performance [5]. This method attempts to achieve the best compromise between the distorted noisy image and the image constructed using sparsity as a prior. The compromise is found by minimizing the MSE and maximizing the sparsity of the coefficients. Since MSE is employed as an optimization criterion, the resulting denoised image might not have the best perceptual quality. This motivated us to replace the role of MSE with SSIM in the framework. To solve this novel optimization problem is not trivial because SSIM is non-convex in nature. There are two key problems that have to be resolved before effective SSIM-based optimization can be done. First, how to optimally decompose an image as a linear combination of basis functions in maximal SSIM, as opposed to minimal MSE sense. Second, how to estimate the best compromise between the noisy and sparse dictionary reconstructed image for maximal SSIM.

We formulate the problem in Section 2 and provide our solutions to issues discussed above in Section 3. Section 4 describes our approach to denoise the images. Simulation results that prove the validity of the proposed approach can be found in Section 5 and finally we conclude in Section 6.

2. PROBLEM FORMULATION

Image denoising in [5] was performed by solving the following optimization problem

$$\begin{aligned}\{\hat{\boldsymbol{\alpha}}_{ij}, \hat{\mathbf{X}}\} = \underset{\boldsymbol{\alpha}_{ij}, \mathbf{X}}{\operatorname{argmin}} \ & \lambda||\mathbf{X}-\mathbf{Y}||_2^2 \\ & + \sum_{ij} \mu_{ij}||\boldsymbol{\alpha}_{ij}||_0 + \sum_{ij} ||\boldsymbol{\Psi}\boldsymbol{\alpha}_{ij} - \mathbf{R}_{ij}\mathbf{X}||_2^2\end{aligned} \tag{1}$$

where $\mathbf{Y}$ is the noisy observed image of size $\sqrt{N} \times \sqrt{N}$, $\mathbf{X}$ is the unknown output denoised image, $\mathbf{R}_{ij}$ is an $n \times N$ matrix that extracts the (ij) block from the image of size $\sqrt{n} \times \sqrt{n}$, $\boldsymbol{\Psi} \in \mathcal{R}^{n \times k}$ is the dictionary with $k > n$ and $\boldsymbol{\alpha}_{ij}$ is the sparse vector of coefficients corresponding to the (ij) block of the image. In Eqn. (1) it is important to note that the first term demands the proximity between the noisy image, $\mathbf{Y}$, and the output image $\mathbf{X}$. However, the second and the third terms make sure that every patch of the output denoised image follow the sparsity prior with bounded error. With the assumption of known dictionary $\boldsymbol{\Psi}$, the two unknowns in Eqn. (1) can be calculated separately by solving the following optimization problems [5].

$$\hat{\boldsymbol{\alpha}}_{ij} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \ \mu_{ij}||\boldsymbol{\alpha}||_0 + ||\boldsymbol{\Psi}\boldsymbol{\alpha} - \mathbf{x}_{ij}||_2^2, \tag{2}$$

where $\mathbf{x}_{ij}$ is (ij) block of the unknown denoised image $\mathbf{X}$ and

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \ \lambda||\mathbf{X}-\mathbf{Y}||_2^2 + \sum_{ij} ||\boldsymbol{\Psi}\boldsymbol{\alpha}_{ij} - \mathbf{R}_{ij}\mathbf{X}||_2^2. \tag{3}$$

To incorporate the SSIM index into the optimization process, Eqns. (2) and (3) are redefined as follows:

$$\hat{\boldsymbol{\alpha}}_{ij} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \ \mu_{ij}||\boldsymbol{\alpha}||_0 + (1 - S(\boldsymbol{\Psi}\boldsymbol{\alpha}, \mathbf{x}_{ij})), \tag{4}$$

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmax}} \;\; S(\mathbf{W}, \mathbf{X}) + \lambda S(\mathbf{X}, \mathbf{Y}), \tag{5}$$

where $S(.,.)$ defines the SSIM measure, $\mathbf{W}$ is the image obtained by averaging the blocks obtained using the sparse coefficients vectors $\hat{\boldsymbol{\alpha}}_{ij}$ calculated by solving optimization problem in Eqn. (4). The use of $1 - S(.,.)$ in Eqn. (4) is also motivated by the fact that it is a variance-normalized $\mathcal{L}_2$ distance [6]. Solution to the optimization problems in Eqns. (4) and (5) is proposed in Section 3.

3. SSIM-BASED APPROXIMATION

This section discusses the solution to the optimization problem in Eqn. (4). Equation (2) can be solved using Orthogonal Matching Pursuit (OMP) [7] by including one atom at a time and stopping when the error $||\boldsymbol{\Psi}\boldsymbol{\alpha}_{ij} - \mathbf{R}_{ij}\mathbf{X}||_2^2$ goes below $T_{mse} = (C\sigma)^2$. C is the noise gain and σ is the standard deviation of the noise. We solve the optimization problem in Eqn. (4) based on the same philosophy. We gather one atom at a time and stop when $SSIM(\boldsymbol{\Psi}\boldsymbol{\alpha}, \mathbf{x}_{ij})$ goes above T_{ssim} which is given as follows.

$$T_{ssim} = 1 - \frac{T_{mse}}{\sigma_{\mathbf{a}}^2 + \sigma_{\mathbf{y}}^2 + C_2}, \tag{6}$$

where C_2 is the constant originally used in SSIM index expression [2] and σ_a^2 is calculated based on current approximation of the block given by $\mathbf{a} := \boldsymbol{\Psi}\boldsymbol{\alpha}$. The set of coefficients $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_k)$ should be calculated such that we get the best approximation $\mathbf{a}$ in terms of SSIM. We search for the stationary points of the partial derivatives of S with respect to $\boldsymbol{\alpha}$. The solution to this problem for orthogonal set of basis is discussed in [6], Here we aim to solve a more general case of linearly independent atoms. The $\mathcal{L}_2$-based optimal coefficients, $\{c_i\}_{i=1}^k$, can be calculated by solving the following system of equations

$$\sum_{j=1}^{k} c_j \langle \psi_i, \psi_j \rangle = \langle \mathbf{y}, \psi_i \rangle, \qquad 1 \leq i \leq k, \tag{7}$$

where $< \cdot, \cdot >$ represents the inner product. The expression for SSIM index is

$$S(\mathbf{a}, \mathbf{y}) = \frac{2\mu_{\mathbf{a}}\mu_{\mathbf{y}} + C_1}{\mu_{\mathbf{a}}^2 + \mu_{\mathbf{y}}^2 + C_1} \frac{2\sigma_{\mathbf{a},\mathbf{y}} + C_2}{\sigma_{\mathbf{a}}^2 + \sigma_{\mathbf{y}}^2 + C_2}, \tag{8}$$

with $\mu_{\mathbf{a}}$ and $\mu_{\mathbf{y}}$ the mean of $\mathbf{a}$ and $\mathbf{y}$ respectively, $\sigma_{\mathbf{a}}$ and $\sigma_{\mathbf{y}}$ the sample variance of $\mathbf{a}$ and $\mathbf{y}$ respectively, and $\sigma_{\mathbf{ay}}$ the covariance of $\mathbf{a}$ and $\mathbf{y}$. The constants C_1 and C_2 are chosen considering the human visual system perception and to ensure numerical stability of the division. First, we write the mean, the variance and the covariance of $\mathbf{a}$ in terms of α with n the size of the current block:

$$\mu_a = \langle \sum_{i=1}^{k} \alpha_i \psi_i \rangle = \sum_{i=1}^{k} \alpha_i \langle \psi_i \rangle, \tag{9}$$

$$\begin{aligned} (n-1)\sigma_{\mathbf{a}}^2 &= \langle \mathbf{a}, \mathbf{a} \rangle - n\langle \mathbf{a} \rangle^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i\alpha_j \langle \psi_i, \psi_j \rangle - n\mu_{\mathbf{a}}^2, \end{aligned} \tag{10}$$

$$\begin{aligned} (n-1)\sigma_{\mathbf{ay}} &= \langle \mathbf{a}, \mathbf{y} \rangle - n\langle \mathbf{a} \rangle \langle \mathbf{y} \rangle \\ &= \sum_{i=1}^{k} \alpha_i \langle \mathbf{y}, \psi_i \rangle - n\mu_{\mathbf{a}}\mu_{\mathbf{y}}, \end{aligned} \tag{11}$$

where $< \cdot >$ represents the sample mean. The partial derivatives are given as follows

$$\frac{\partial \mu_{\mathbf{a}}}{\partial \alpha_i} = \langle \psi_i \rangle, \tag{12}$$

$$(n-1)\frac{\partial \sigma_{\mathbf{a}}^2}{\partial \alpha_i} = 2\sum_{j=1}^{k} \alpha_j \langle \psi_i, \psi_j \rangle - 2n\mu_{\mathbf{a}} \langle \psi_i \rangle, \tag{13}$$

$$(n-1)\frac{\partial \sigma_{\mathbf{ay}}}{\partial \alpha_i} = \langle \mathbf{y}, \psi_i \rangle - n\mu_{\boldsymbol{y}} \langle \psi_i \rangle, \tag{14}$$

After subtracting the corresponding DC values from all the blocks in the image, we are interested only in the particular case where the atoms are made of oscillatory function, i.e. when $\langle \psi_i \rangle = 0$, for $1 \leq i \leq k$. From logarithmic differentiation of Eqn. (8) combined with Eqns. (12)-(14), we have

$$\frac{1}{S}\frac{\partial S}{\partial \alpha_i} = \frac{2\langle \mathbf{y}, \psi_i \rangle}{(n-1)2\sigma_{\mathbf{a},\mathbf{y}} + C_2} - \frac{2\left(\sum_{j=1}^{k} \alpha_j \langle \psi_i, \psi_j \rangle\right)}{(n-1)\left(\sigma_{\mathbf{a}}^2 + \sigma_{\mathbf{y}}^2 + C_2\right)} \tag{15}$$

We equate (15) to zero in order to find the stationary points. The result is the following linear system of equations

$$\sum_{j=1}^{k} \alpha_j \langle \psi_i, \psi_j \rangle = \beta \langle \mathbf{y}, \psi_i \rangle, \qquad 1 \leq i \leq k, \tag{16}$$

where

$$\beta = \frac{\sigma_{\mathbf{a}}^2 + \sigma_{\mathbf{y}}^2 + C_2}{2\sigma_{\mathbf{ay}} + C_2}, \tag{17}$$

Comparing $\boldsymbol{\alpha}$ with the optimal coefficients in $\mathcal{L}_2$ sense denoted by $\mathbf{c}$ and given by Eqn. (7) results in the following solution:

$$\alpha_i = \beta c_i, \qquad 1 \leq i \leq k, \tag{18}$$

which implies that the optimal SSIM-based solution is just a scaling of the optimal $\mathcal{L}_2$-based solution. The last step is to find β. After substituting the value for α_i in the expression for β and then isolating for β gives us the following quadratic equation

$$\beta^2(B - A) + \beta C_2 - \sigma_{\mathbf{y}}^2 - C_2 = 0, \tag{19}$$

where

$$A = \frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{k} c_i c_j \langle \psi_i, \psi_j \rangle, \tag{20}$$

$$B = \frac{2}{n-1}\sum_{j=1}^{k} c_j \langle \mathbf{y}, \psi_j \rangle. \tag{21}$$

Solving for β and picking a positive value for maximal SSIM gives us

$$\beta = \frac{-C_2 + \sqrt{C_2^2 + 4(B - A)(\sigma_{\mathbf{y}}^2 + C_2)}}{2(B - A)}. \tag{22}$$

The solution to the optimization problem in Eqn. (4) is complete and we have the coefficients $\boldsymbol{\alpha}_{ij}$ to solve the optimization problem in Eqn. (5). We use gradient-descent approach to solve the optimization problem in Eqn. (5). The gradient of SSIM expression in Eqn. (8) is given as follows [8]

$$\begin{aligned} \vec{\nabla}_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y}) = \frac{2}{nB_1^2 B_2^2}[&A_1 B_1 (B_2 \mathbf{x} - A_2 \mathbf{y}) \\ &+ B_1 B_2 (A_2 - A_1)\mu_{\mathbf{x}}\mathbf{1} + A_1 A_2 (B_1 - B_2)\mu_{\mathbf{y}}\mathbf{1}], \end{aligned} \tag{23}$$

Table 1. SSIM and PSNR comparisons of image denoising results

Image	Barbara				Lena				Peppers				House			
Noise std	20	25	50	100	20	25	50	100	20	25	50	100	20	25	50	100
	PSNR comparison (in dB)															
Noisy	22.11	20.17	14.15	8.13	22.11	20.17	14.15	8.13	22.11	20.17	14.15	8.13	22.11	20.17	14.15	8.13
K-SVD	30.85	29.55	25.44	21.65	32.38	31.32	27.79	24.46	30.80	29.72	26.10	21.84	33.16	32.12	28.08	23.54
Proposed	30.88	29.53	25.50	21.74	32.26	31.28	27.80	24.53	30.84	29.84	26.25	21.98	33.04	32.09	28.13	23.59
	SSIM comparison															
Noisy	0.593	0.503	0.241	0.084	0.531	0.443	0.204	0.074	0.529	0.442	0.212	0.076	0.452	0.368	0.166	0.057
K-SVD	0.894	0.859	0.708	0.519	0.903	0.877	0.733	0.550	0.905	0.883	0.782	0.601	0.909	0.890	0.779	0.549
Proposed	0.906	0.875	0.733	0.526	0.913	0.888	0.754	0.573	0.913	0.894	0.797	0.627	0.915	0.901	0.795	0.574

$$A_1 = 2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + C_1, \qquad A_2 = 2\sigma_{\mathbf{xy}} + C_2,$$

$$B_1 = \mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + C_1, \qquad B_2 = \sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + C_2,$$

where $\mu_{\mathbf{x}}$, $\sigma_{\mathbf{x}}^2$ and $\sigma_{\mathbf{xy}}$ represent, respectively, the sample mean of the components of $\mathbf{x}$, the sample variance of $\mathbf{x}$, and the sample covariance of $\mathbf{x}$ and $\mathbf{y}$, and n is the number of pixels in the local image patch. We follow the gradient for few steps until converging to a stationary point. We initialize $\hat{\mathbf{X}}$ as the best MSE solution.

4. IMAGE DENOISING

The proposed image denoising algorithm is summarized as follows

Algorithm 1: SSIM-inspired image denoising

1. *Initialize:* $\mathbf{X} = \mathbf{Y}$, $\mathbf{\Psi} =$ overcomplete DCT dictionary
2. *Repeat J times*
   - *Sparse coding stage:* use SSIM-optimal OMP to compute the representation vectors α_{ij} for each patch
   - *Dictionary update stage:* Use K-SVD [9] to calculate the updated dictionary and coefficients. Calculate SSIM-optimal coefficients using Eqns. (18) and (22)
3. *Global Reconstruction:* Use gradient descent algorithm to optimize Eqn. (5), where the SSIM gradient is given by Eqn. (23).

The modified OMP pursuit algorithm is explained in Algorithm 2. There are two main differences between the OMP algorithm [7] and the one proposed in this work. First, the stopping criterion is based on SSIM. The shapes of MSE and SSIM contours are different and are explained in [2] in detail. Defining the stopping criterion according to SSIM essentially means that we are modifying the set of *accepted* points (image patches) around the noisy image patch which can be represented as the linear combination of dictionary atoms. This way we are omitting image patches in the direction of structural distortion and including the ones which are in the same direction as the original image patch in the set of *acceptable* image patches. Therefore, we can expect to see more structures in the image constructed using sparsity as a prior. Second, we calculate the SSIM-optimal coefficients from the optimal coefficients in $\mathcal{L}_2$-sense using the derivation in Section 3, which are scalar multiple of the optimal $\mathcal{L}_2$-based coefficients.

Algorithm 2: SSIM-inspired Orthogonal Matching Pursuit

Initialize: $\mathbf{D} = \{\}$ set of selected atoms, $S_{opt} = 0$, $\mathbf{r} = \mathbf{Y}$
while $S_{opt} < T_{ssim}$

- Add the next best atom in $\mathcal{L}_2$ sense to $\mathbf{D}$
- Find the optimal $\mathcal{L}_2$-based coefficient(s) using Eqn. (7)
- Find the optimal SSIM-based coefficient(s) using Eqn. (18)
- Update the residual $\mathbf{r}$
- Find SSIM-based approximation $\mathbf{a}$
- Calculate $S_{opt} = S(\mathbf{a}, \mathbf{y})$

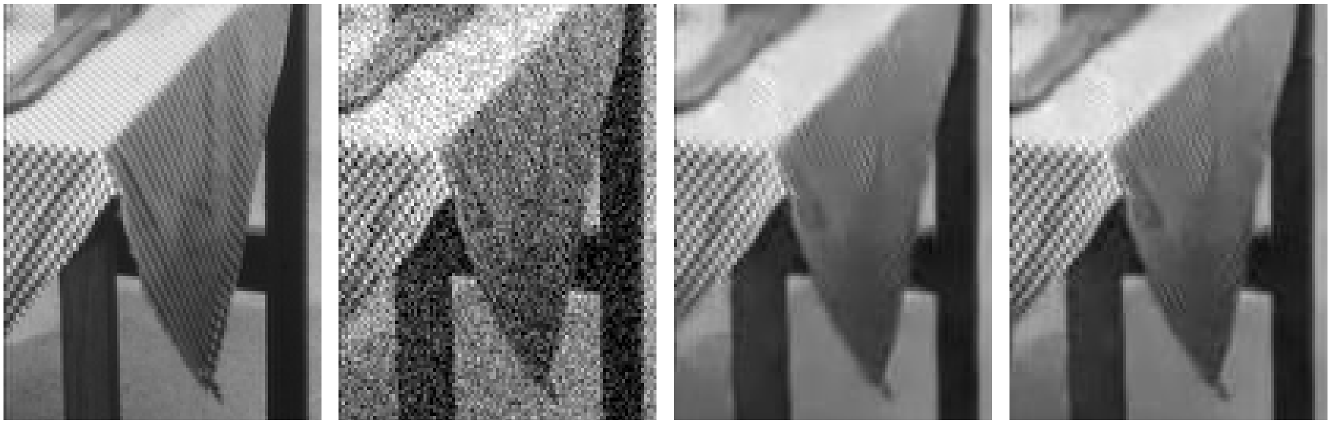
end

5. SIMULATION RESULTS

The proposed image denoising scheme is tested on various images with different amount of noise. In all the experiments, the dictionary used were of size 64×256, designed to handle patches of 8×8 pixels. The value of noise gain, C , is selected to be 1.15 and $\lambda = 30/\sigma$ [5]. Table 1 shows the results for images *Barbara, Lena, Peppers, House*. It also compares the K-SVD method [5] with the proposed denoising method. It can be observed that the proposed denoising method achieves better performance in terms of SSIM which implies better perceptual quality of the denoised image. Figure 1 shows that the denoised images using K-SVD [5] and the proposed methods. It can be seen that the proposed denoising scheme preserves the structures better and therefore has better perceptual image quality.

6. CONCLUSIONS

In this paper, we attempt to combine perceptual image fidelity measurement with optimal sparse signal representation in the context of image denoising. We proposed an algorithm to solve for the optimal coefficients for sparse and redundant dictionary in maximal SSIM sense. We also developed a gradient descent approach to achieve the best compromise between the distorted noisy image and the image reconstructed using sparse representation. Our simulations demonstrate promising results and also indicate the potential of SSIM to replace the ubiquitous PSNR/MSE as the optimization criterion in image processing applications.



(a) Original image (b) Noisy image ($\sigma = 25$) (c) Denoised by K-SVD (d) Denoised by proposed method

(e) Original image (f) Noisy image ($\sigma = 25$) (g) Denoised by K-SVD (h) Denoised by proposed method

Fig. 1. Visual comparison of denoising results

7. ACKNOWLEDGMENT

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada in the forms of Discovery, Strategic and CRD Grants, and in part by Ontario Early Researcher Award, which are gratefully acknowledged.

8. REFERENCES

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

[3] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R.W. Heath, "Design of linear equalizers optimized for the structural similarity index," *IEEE Transactions on Image Processing*, vol. 17, no. 6, pp. 857 –872, jun. 2008.

[4] Z. Wang, Q. Li, and X. Shang, "Perceptual image coding based on a maximum of minimal structural similarity criterion," *IEEE International Conference on Image Processing*, vol. 2, pp. II –121 –II –124, sep. 2007.

[5] 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, 2006.

[6] D. Brunet, E. R. Vrscay, and Z. Wang, "Structural similarity-based approximation of signals and images using orthogonal bases," in *International Conference on Image Analysis and Recognition*, Portugal, June 2010.

[7] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Twenty Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 40–44, 1993.

[8] Z. Wang and E. P. Simoncelli, "Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities," *Journal of Vision*, vol. 8, no. 12, 2008.

[9] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.