A WEIGHTED DISCRIMINATIVE APPROACH FOR IMAGE DENOISING WITH OVERCOMPLETE REPRESENTATIONS

Amir Adler[1]*, Yacov Hel-Or*[2] *and Michael Elad*[1]

[1] Computer Science Department – The Technion, Haifa 32000, Israel
[2] Efi Arazi School of Computer Science, The Interdisciplinary Center, Herzlia 46150, Israel

ABSTRACT

We present a novel weighted approach for shrinkage functions learning in image denoising. The proposed approach optimizes the shape of the shrinkage functions and maximizes denoising performance by emphasizing the contribution of sparse overcomplete representation components. In contrast to previous work, we apply the weights in the overcomplete domain and formulate the restored image as a weighted combination of the post-shrinkage overcomplete representations. We further utilize this formulation in an offline Least Squares learning stage of the shrinkage functions, thus adapting their shape to the weighting process. The denoised image is reconstructed with the learned weighted shrinkage functions. Computer simulations demonstrate superior shrinkage-based denoising performance.

Index Terms— denoising, shrinkage, weight, sparsity.

1. INTRODUCTION

Image denoising in the overcomplete domain has driven substantial research efforts since the pioneering work of Coifman and Donoho [1]. Most of the research has been focused in developing new overcomplete representations and optimizing the shrinkage mapping functions (MFs). Commonly, the conventional techniques design the MFs using a "descriptive" approach based on the statistical prior of the transform domain. In this paper we consider the utilization of local sparsity information, available in the overcomplete domain, for the performance enhancement of a "discriminative" shrinkage restoration approach [2]. In the discriminative approach, the MFs are learned from example images, targeting the denoising performance directly, rather than a descriptive approach which focuses on determining the statistical prior. A weighted approach for image denoising in overcomplete domains was proposed in [3], where a per-transform and per-pixel sparsity based weight function was utilized in conjunction with a hard-thresholding scheme. The derivation of the weight function was performed regardless of the shrinkage function, thus not utilizing the weight information during the shaping process of the shrinkage functions.

In this paper we propose to utilize a weighted approach in two stages as follows a) the weight function is applied during the discriminative learning process of the shrinkage functions, thus adapting the shrinkage functions to the weighted reconstruction stage b) The weights are applied in the overcomplete domain during the denoising process. This strategy enables the design of reduced complexity restoration schemes that are based on convolutions with basis kernels.

2. PROBLEM FORMULATION

We assume the following image degradation model

$$\mathbf{y} = \mathbf{x} + \mathbf{n}. \tag{1}$$

where $\mathbf{x} \in \Re^{N \times 1}$ is a clean image (column stacked) contaminated with additive white Gaussian noise $\mathbf{n} \sim N(0, \sigma I) \in \Re^{N \times 1}$ and $\mathbf{y} \in \Re^{N \times 1}$ is the noisy image. Shrinkage-based denoising is based on sparse decompositions with linear transforms, where low magnitude coefficients have low SNR. As shown in [1], a scalar thresholding (i.e. shrinkage) function applied to every component of the decomposition improves the SNR performance of the post-shrinkage reconstructed signal. In the classical translation invariant Wavelet shrinkage approach, a set of K fully decimated transforms are applied to the noisy image. Each one of the transforms is a different shifted version of the un-shifted "mother" transform. In our approach, the overcomplete representation constitutes of K sub-bands generated by filtering the image with the respective basis kernels of each sub-band. It can be easily proved [2] that when the Wavelet transform consists of windowed basis functions (e.g. DCT) the two approaches are equivalent.

Given is an overcomplete linear transform with K sub-bands

$$B = \left[B_1^T, B_2^T, \cdots, B_K^T \right]^T \in \Re^{NK \times N}.$$

where each matrix $B_i \in \Re^{N \times N}$ is the (block-circulant) convolution matrix of the respective basis kernel. In the rest of this paper we will utilize an un-decimated $L \times L$ windowed DCT (UDCT) as an overcomplete transform example. For this case, the representation redundancy factor is $K = L^2$ since the transform is defined to include all possible DCT window shifts. The matrices B_i are derived from the basis kernels of the DCT sub-bands.

The overcomplete decomposition of the noisy image is given by

$$\mathbf{y}_B = \begin{bmatrix} \mathbf{y}_{B_1} \\ \mathbf{y}_{B_2} \\ \vdots \\ \mathbf{y}_{B_K} \end{bmatrix} = \begin{bmatrix} B_1 \mathbf{y} \\ B_2 \mathbf{y} \\ \vdots \\ B_K \mathbf{y} \end{bmatrix} = B\mathbf{y}. \tag{2}$$

The shrinkage operation is performed by a scalar MF $\Psi_k\left(\mathbf{y}_{B_k}\right)$ (operating element-wise) per sub-band, and the reconstructed image is given by

$$\hat{\mathbf{x}}(\mathbf{y}) = \left(\sum_{k=1}^{K} B_k^T B_k\right)^{-1} \sum_{k=1}^{K} B_k^T \Psi_k\left(\mathbf{y}_{B_k}\right). \qquad (3)$$

The last equation is a straightforward Least Squares (LS) solution of the linear set of equations:

$$B_k \mathbf{x} = \Psi_k\left(\mathbf{y}_{B_k}\right), \qquad k = 1, 2, .., K. \qquad (4)$$

using pseudo-inverse. In the discriminative approach [2] the denoising stage (3) is preceded by the MFs learning stage. The MFs are approximated by a piece-wise linear model and jointly estimated using set of example images whose clean and noisy counterparts are given offline. The piece-wise linear approximation is performed with the Slice Transform (see Appendix A)

$$\Psi_k\left(\mathbf{y}_{B_k}\right) \approx S_{\mathbf{q}_k}\left(\mathbf{y}_{B_k}\right)\mathbf{p}_k. \qquad (5)$$

The vector $\mathbf{p}_k$ controls the shape of the k-th MF and it is the design parameter per sub-band.

The Slice Transform (SLT) approximation gives rise to a simple LS formulation of the MFs learning process

$$\hat{\mathbf{p}}\left(\mathbf{x}^e, \mathbf{y}^e\right) = \arg\min_{\mathbf{p}} \left\|\hat{\mathbf{x}}\left(\mathbf{y}^e, \mathbf{p}\right) - \mathbf{x}^e\right\|^2. \qquad (6)$$

Where $\mathbf{x}^e$ and $\mathbf{y}^e$ are the clean and noisy example images, respectively, and $\mathbf{p} = \left[\mathbf{p}_1^T, \mathbf{p}_2^T, \cdots, \mathbf{p}_K^T\right]^T$.

In this paper we augment this approach by utilizing sparsity information, extracted from the overcomplete domain and represented by the weight vector $\mathbf{w} = \left[w_1, w_2, \cdots, w_N\right]^T$.

We are therefore interested in solving the following problems:

1. Learning the MFs given the example images and the respective weights, namely deriving $\hat{\mathbf{p}}\left(\mathbf{x}^e, \mathbf{y}^e, \mathbf{w}\right)$.

2. Estimating the clean image given the noisy image, the respective weights and the learned MFs, namely deriving $\hat{\mathbf{x}}\left(\mathbf{y}, \mathbf{w}, \mathbf{p}\right)$.

3. Defining effectively the weight function.

3. WEIGHTED RESTORATION IN OVER-COMPLETE DOMAINS

By applying the shrinkage operation to the noisy image and weighting all sub-bands of the noisy and estimated images, it is expected that the following linear set of equations hold:

$$WB_k \mathbf{x} = W\Psi_k\left(\mathbf{y}_{B_k}\right), \qquad k = 1, 2, .., K. \qquad (7)$$

where

$$W = \begin{pmatrix} w_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & w_N \end{pmatrix} \in R^{N \times N}.$$

and we assume a unique weight value per image location (pixel), as explained in section 3.2. The reconstructed image is given by the pseudo-inverse

$$\hat{\mathbf{x}}(\mathbf{y}, \mathbf{w}) = W_\pi^{-1} \sum_{k=1}^{K} B_k^T W^2 \Psi_k\left(\mathbf{y}_{B_k}\right). \qquad (8)$$

where

$$W_\pi = \sum_{k=1}^{K} B_k^T W^2 B_k.$$

In the case where B is a tight frame and an $L \times L$ block based transform (such as the $L \times L$ UDCT), the matrix W_π can be shown to be

$$W_\pi = \begin{pmatrix} w_{\pi,1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & w_{\pi,N} \end{pmatrix}.$$

where $diag\left(W_\pi\right) = \mathbf{w}_\pi = H_\pi \mathbf{w}$ and H_π is the (block-circulant) convolution matrix of the kernel

$$\pi = \frac{1}{L^2}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathfrak{R}^{L \times L}.$$

The proof of this derivation is beyond the scope of this paper.

3.1. Weighted Learning of the Shrinkage Functions

Considering the weighted restoration scheme in the overcomplete domain (8), our goal is to estimate the shrinkage MF of every sub-band in a process that would take into account the weighting operation. By utilizing the SLT approximation (5) we obtain the following weighted reconstruction for the noisy example image

$$\begin{aligned} \hat{\mathbf{x}}\left(\mathbf{y}^e, \mathbf{w}, \mathbf{p}\right) &= W_\pi^{-1} \sum_{k=1}^{K} B_k^T W^2 S_{\mathbf{q}_k}\left(\mathbf{y}_{B_k}\right)\mathbf{p}_k \\ &= W_\pi^{-1} L_W\left(\mathbf{y}_B\right)\mathbf{p}. \end{aligned} \qquad (9)$$

where

$$L_W\left(\mathbf{y}_B\right) = \left[H_1, H_2, \cdots, H_K\right]$$

and

$$H_i = B_i^T W^2 S_{\mathbf{q}_i}\left(y_{B_i}\right).$$

The MFs estimation process is performed offline utilizing example images. We formulate the learning process as a LS estimation problem

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \left\|\hat{\mathbf{x}}\left(\mathbf{y}^e, \mathbf{w}, \mathbf{p}\right) - \mathbf{x}^e\right\|^2. \qquad (10)$$

Substituting (9) into (10) results in the closed form solution

$$\hat{\mathbf{p}}\left(\mathbf{x}^e, \mathbf{y}^e, \mathbf{w}\right) = \left(\tilde{L}_W^T \tilde{L}_W\right)^{-1} \tilde{L}_W^T \mathbf{x}^e. \qquad (11)$$

where

$$\tilde{L}_W = W_\pi^{-1} L_W\left(\mathbf{y}_B^e\right).$$

3.2. The Weight Function

The weight we use is a function of the Sparsity Concentration Index (SCI) [4] of a vector $\mathbf{a} \in \mathfrak{R}^{M \times 1}$ given by:

$$SCI(\mathbf{a}) = \left(M \cdot \max_i |a_i| / \|\mathbf{a}\|_1 - 1\right) \Big/ (M-1) \in [0,1]. \qquad (12)$$

Note, that $SCI(\mathbf{a}) = 1$ if $\mathbf{a}$ has a single non-zero element and $SCI(\mathbf{a}) = 0$ if all the elements are equal. The weight function is given by

$$w(\mathbf{a}) = \rho_{SCI}(\mathbf{a}) = \big(SCI(\mathbf{a})\big)^s. \qquad (13)$$

where we have found that powers in the range $s \in [1,2]$ provide good performance. The choice of the SCI metric is motivated by the following example: assume that B is a $L \times L$ UDCT transform and we are interested in denoising a certain pixel in a noisy image. The pixel of interest is contained in L^2 possible overlapping DCT windows. Some of these provide a highly sparse representation, since they contain a relatively simple structure, whereas others may be less sparse. By strengthening the weights of the sparser windows, we enhance the reliability of the denoising.

The computation of the weights is performed as follows: weight value for a pixel location, i, is computed from the DCT coefficients of a window centered at location i. Since there are N different shifted windows, there are N different weight values (one per window location). Given an overcomplete representation $\mathbf{z}_{B_k} = B_k \mathbf{z}, \, k = 1..K$ of an image $\mathbf{z}$, the i-th element of the weight vector $\mathbf{w}$ is given by

$$w_i(\mathbf{z}) = \rho_{SCI}\left(\big[z_{B_1}(i), z_{B_2}(i), \cdots, z_{B_K}(i)\big]^T\right), \quad i = 1..N. \qquad (14)$$

where $z_{B_k}(j)$ is the j-th element of the vector $\mathbf{z}_{B_k}$.

3.3. The proposed Restoration Algorithm

The restoration algorithm includes two stages, summarized in Table I. During the first stage the MFs are estimated from the set of example images. The corresponding weights are computed from the clean image, thus, extracting the most accurate sparsity information. The estimated MFs are represented by the parameter vector $\hat{\mathbf{p}}$. In the second stage, a noisy image is denoised utilizing the learned MFs and the weights. Since the clean image is not available at this stage, the weights are calculated from the post-shrinkage overcomplete representation of the noisy image. Weights calculation in this approach was verified to be more accurate, due to the reduction of noisy components, compared to calculation from the pre-shrinkage overcomplete representation.

4. PERFORMANCE EVALUATION

We performed computer simulations of the proposed approach and compared it with the un-weighted approach [2]. The tested images were taken from the collection of [5] and every denoising result was averaged over 10 noise realizations. A 9x9 UDCT was chosen as the overcomplete representation transform. The shrinkage MFs were learned with a natural image that was verified to provide good denoising results. Table II summarizes PSNR results which verify a consistent advantage of the weighed approach over the un-weighted approach. Fig. 1 presents PSNR improvement over the un-weighted results as a function of the noise level: it can be seen that the weighted approach provides an advantage that increases with the noise level. Fig. 2 presents examples of several learned MFs: it can be seen that the weighting impacts mostly the boosting portion of mid-amplitude transform coefficients. Fig. 3 compares the visual quality of denoising the image Peppers (best viewed in the electronic version of this paper), with noise level $\sigma = 25$. It can be seen that the weighted approach results in smoother edges, compared to the un-weighted approach.

Learning Stage (offline)

Input: example images $\mathbf{x}^e$, $\mathbf{y}^e$.

Output: learned MFs of all sub-bands represented by $\hat{\mathbf{p}}$.

1. Compute the overcomplete representations
$$\mathbf{x}^e_{B_k} = B_k \mathbf{x}^e, \, \mathbf{y}^e_{B_k} = B_k \mathbf{y}^e.$$

2. Compute the weights $w_i(\mathbf{x}^e)$, $i = 1..N$:
$$w_i(\mathbf{x}^e) = \rho_{SCI}\left(\big[x^e_{B_1}(i), x^e_{B_2}(i), \cdots, x^e_{B_K}(i)\big]^T\right).$$

3. Estimate the MFs according to:
$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \left\| \hat{\mathbf{x}}(\mathbf{y}^e, \mathbf{w}, \mathbf{p}) - \mathbf{x}^e \right\|^2.$$

Restoration Stage (online)

Input: noisy image $\mathbf{y}$ and MFs represented by $\hat{\mathbf{p}}$.

Output: denoised image $\hat{\mathbf{x}}$.

1. Compute the overcomplete representation $\mathbf{y}_{B_k} = B_k \mathbf{y}$

2. Apply the shrinkage functions and obtain $\tilde{\mathbf{y}}_{B_k} = \Psi_k(\mathbf{y}_{B_k})$

3. Compute the weights vector $w_i(\mathbf{y})$, $i = 1..N$:
$$w_i(\mathbf{y}) = \rho_{SCI}\left(\big[\tilde{y}_{B_1}(i), \tilde{y}_{B_2}(i), \cdots, \tilde{y}_{B_K}(i)\big]^T\right).$$

4. Reconstruct the image according to
$$\hat{\mathbf{x}} = W_\pi^{-1} \sum_{k=1}^{K} B_k^T W^2 \Psi_k(\mathbf{y}_{B_k}).$$

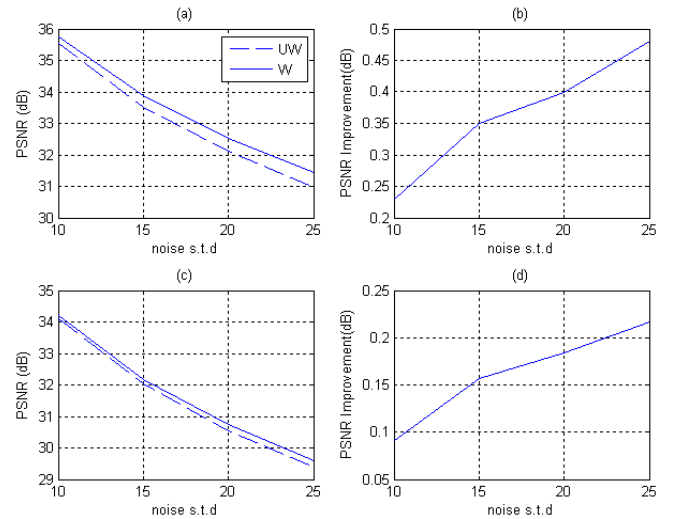**Table I**: The Proposed Restoration Algorithm

Fig. 1: Weighted (W) vs. un-weighted (UW) results: (a) PSNR of image House (b) PSNR improvement of House (c) averaged PSNR over all tested images in the collection of [5] (d) averaged PSNR improvement over all tested images in the collection of [5]

Image		Barb.	Boat	Fgrpt.	House	Lena	Pepp.
σ=10	UW	34.19	33.49	32.27	35.53	35.37	33.84
	W	**34.20**	**33.52**	32.27	**35.76**	**35.44**	**34.05**
σ=15	UW	31.95	31.55	29.94	33.52	33.47	31.73
	W	**31.97**	**31.64**	**29.98**	**33.87**	**33.64**	**32.00**
σ =20	UW	30.36	30.19	28.36	32.11	32.10	30.20
	W	**30.41**	**30.29**	**28.42**	**32.51**	**32.29**	**30.50**
σ =25	UW	29.09	29.11	27.15	30.95	31.02	29.04
	W	**29.19**	**29.22**	**27.22**	**31.43**	**31.19**	**29.41**

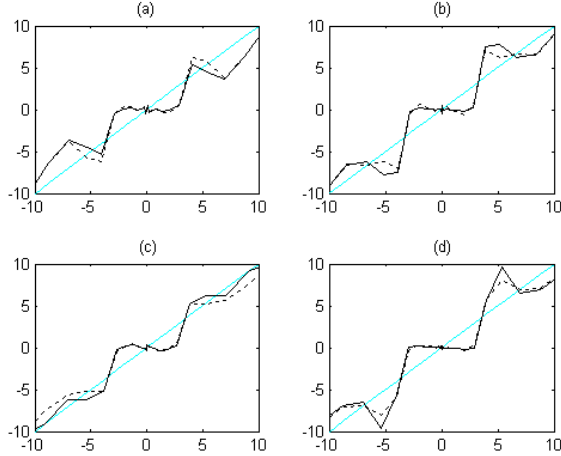**Table II**: Denoising PSNR (dB) results: weighted (W) vs. un-weighted (UW)

Fig. 2: Weighted (solid) vs. un-weighted (dashed) learned MFs: (a) sub-band #17 (b) sub-band #25 (c) sub-band #47 (d) sub-band #65
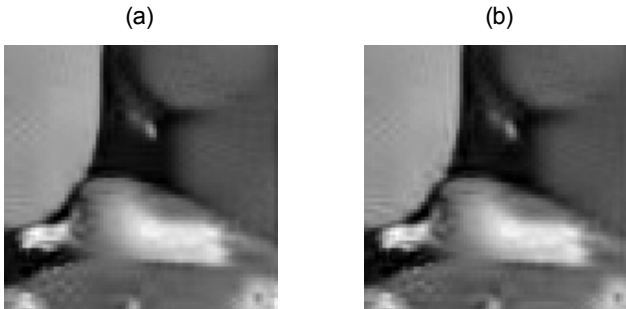

Fig. 3: Denoising of Peppers: (a) weighted (b) un-weighted

5. CONCLUSIONS

We have introduced the application of sparsity-induced weights to learn an optimal set of MFs for weighted shrinkage image denoising. We formulated the denoised image as a weighted combination of the post-shrinkage transform coefficients. Using this formulation, we expressed the MFs learning process as a LS estimation problem from example images. Computer simulations demonstrated superior shrinkage-based denoising performance. The improvement in PSNR was further shown to increase with the noise level.

APPENDIX A: THE SLICE TRANSFORM

Let $x \in [a,b] \in \Re$ be a real value. The interval $[a,b]$ is divided into M bins whose boundaries form a vector $\mathbf{q} = [q_0, q_1, \cdots, q_M]^T$ such that $a = q_0 < q_1 \cdots < q_M = b$. The value x is associated with a single bin $\pi(x) \in \{1 \cdots M\}$ and a corresponding residue $r(x)$, where $\pi(x) = j$ if $x \in [q_{j-1}, q_j)$ and $r(x) = (x - q_{\pi(x)-1})/(q_{\pi(x)} - q_{\pi(x)-1}) \in [0,1)$. The value x can be expressed as a linear combination of $q_{\pi(x)}$ and $q_{\pi(x)-1}$

$$x = r(x)q_{\pi(x)} + (1 - r(x))q_{\pi(x)-1}.$$

Assume now that $\mathbf{x}$ is *N*-dimensional vector whose elements satisfy $x_i \in [a,b]$. The SLT of $\mathbf{x}$ is given by $\mathbf{x} = S_{\mathbf{q}}(\mathbf{x})\mathbf{q}$ where the matrix $S_{\mathbf{q}}(\mathbf{x}) \in \Re^{Nx(M+1)}$ is defined as follows

$$\left[S_{\mathbf{q}}(\mathbf{x}) \right]_{(i,j)} = \begin{cases} r(x_i), & if\ \pi(x_i) = j \\ 1 - r(x_i), & f\ \pi(x_i) = j+1 \\ 0, & otherwise \end{cases}.$$

According to [2], substituting the boundary vector $\mathbf{q}$ with a different vector $\mathbf{p}$ in the form of $M_{\mathbf{q},\mathbf{p}}\{\mathbf{x}\} = S_{\mathbf{q}}(\mathbf{x})\mathbf{p}$ performs a piecewise linear mapping of the values in $\mathbf{x}$ such that the values $\{x \in [q_{j-1}, q_j)\}$ are mapped linearly to the interval $[p_{j-1}, p_j)$. The substitution property is the key principle behind the approximation of the shrinkage MFs.

7. REFERENCES

[1] R. Coifman and D. Donoho, "Translation invariant denoising," in *Lecture Notes in Statistics: Wavelets and Statistics*, New York: Springer-Verlag, pp. 125--150, 1995.

[2] Y. Hel-Or and D. Shaked, "A Discriminative approach for Wavelet Shrinkage Denoising", *IEEE Trans. on Image Processing*, vol. 17, no. 4, April 2008.

[3] O. G. Guleryuz, "Weighted Averaging for Denoising With Overcomplete Dictionaries ", *IEEE Trans. on Image Processing*, vol. 16, no. 12, December 2007.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, February 2009.

[5] BLS-GSM Test Images, available online: http://decsai.ugr.es/~javier/denoise/test_images/index.htm