A Shrinkage Learning Approach for Single Image Super-Resolution with Overcomplete Representations

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

[1] Computer Science Department, The Technion, Haifa, Israel
[2] Efi Arazi School of Computer Science,
The Interdisciplinary Center, Herzelia, Israel

Abstract. We present a novel approach for online shrinkage functions learning in single image super-resolution. The proposed approach leverages the classical *Wavelet Shrinkage* denoising technique where a set of scalar shrinkage functions is applied to the wavelet coefficients of a noisy image. In the proposed approach, a unique set of learned shrinkage functions is applied to the overcomplete representation coefficients of the interpolated input image. The super-resolution image is reconstructed from the post-shrinkage coefficients. During the learning stage, the low-resolution input image is treated as a reference high-resolution image and a super-resolution reconstruction process is applied to a scaled-down version of it. The shapes of all shrinkage functions are jointly learned by solving a Least Squares optimization problem that minimizes the sum of squared errors between the reference image and its super-resolution approximation. Computer simulations demonstrate superior performance compared to state-of-the-art results.

1 Introduction

Single Image Super-Resolution (SISR) is the process of reconstructing a high-resolution image from an observed low-resolution image. Typical applications include zoom-in of still images in digital cameras, scaling-up an image before printing and conversion from low-definition to high-definition video. SISR is an inverse problem, associated with the following linear degradation model

$$\mathbf{y} = DH\mathbf{x}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^n$ is the observed low-resolution input image (column-stacked), $\mathbf{x} \in \mathbb{R}^{nL}$ is the unknown high-resolution image, $H \in \mathbb{R}^{nL \times nL}$ is a blurring filter (block-circulant) convolution matrix and $D \in \mathbb{R}^{n \times nL}$ is a down-sample operator matrix, decimating the image by a factor of $\sqrt{L}$ along the horizontal and vertical dimensions.

A solution to the SISR problem is an approximation $\hat{\mathbf{x}}$ to the unknown high-resolution image $\mathbf{x}$. Since the linear system (1) is underdetermined, there are infinitely many solutions $\hat{\mathbf{x}}$ that can "explain" the observed image $\mathbf{y}$. For this

reason, there are various approaches addressing the SISR problem. The simplest techniques are the bi-linear and bi-cubic interpolators. These interpolators utilize a polynomial approximation model to compute each missing pixel from a small local neighborhood of it, often generating blurry results and stair-case shaped edges.

State-of-the-art SISR reconstruction is based on a sparse-representation approach [1], [2] where a set of high-resolution and low-resolution dictionaries are learned from example images. In this approach, a sparse coding process is applied to small overlapping patches, extracted in a raster-scan order from the observed image. The sparse representation coefficients (i.e the outcome of the sparse coding process) of each low-resolution patch are assumed to faithfully represent each corresponding (unknown) high-resolution patch by replacing the low-resolution dictionary with its high-resolution counterpart. The super-resolution image is reconstructed by fusion of all of the overlapping high-resolution patches. A similar approach was proposed in [3], where a reduced redundancy dictionary was employed to accelerate the SISR process. The sparse-representation approach evolved from an example-based approach [4], where a dictionary of $100,000$ pairs of low-resolution and high-resolution images patches was utilized in conjunction with a markov-network model to search-and-match the corresponding high-resolution patches. A combination of the example-based approach with multi-frame super-resolution was proposed in [5], where patch repetitions within an image were exploited in a multi-scale approach. Additional example-based approaches such as learning the prior parameters, learning the posterior and building example-based regularization expression are reviewed in [6]. A shrinkage-based approach was introduced in [7] where a hard-thresholding function was iteratively applied to DCT transform coefficients. This approach was later augmented in [8], where the Contourlet transform was chosen as the overcomplete transform.

We propose to extend the shrinkage based approach and employ online-learned shrinkage functions with an overcomplete representation. The proposed approach leverages the *discriminative* learning technique suggested in [9] for wavelet denoising. In the discriminative approach, the shapes of all shrinkage functions are learned offline from example images (rather then learning the parameters of a probability distribution model of the transform coefficients). In the proposed approach, we apply the discriminative approach to the SISR problem and exploit the scale-invariant property of natural images [10] to learn the shrinkage functions directly from the input image.

Contributions. The contributions presented in this paper are two-fold: 1) Introduction of the learned shrinkage approach [9] to solve the SISR problem, in contrast to the hard-thresholding approach previously introduced for SISR in [8], [7]. 2) Introduction of the online learning approach, where the shrinkage functions are learned directly from the observed input image - in contrast to the offline example-based approach as suggested in [9]. The advantage of the online approach is that online-learned shrinkage functions capture the statistical properties of the observed image (to be scaled-up), rather than the statistical

properties of other images. Performance evaluation of the proposed approach demonstrate superior performance compared to the sparse representation[1] state-of-the-art approach [1], [2].

This paper is organized as follows: Section 2 presents shrinkage-based restoration theory for the unitary and over-complete cases. Section 3 describes the Slice Transform (SLT) which is a piece-wise linear model utilized for the representation and learning of the shrinkage functions. Section 4 presents the proposed SISR algorithm concept along with a detailed explanation of the shrinkage-learning stage and the super-resolution reconstruction stage. Section 5 overviews performance evaluation along with a comparison versus the state-of-the-art approach.

2   Shrinkage-Based Image Restoration

This section provides an overview of shrinkage-based image restoration in the unitary and overcomplete cases. The discussion evolves from an image denoising problem and the connection to the SISR problem is established in the last subsection. Consider the following image degradation model,

$$\mathbf{v} = \mathbf{u} + \mathbf{m}, \tag{2}$$

where $\mathbf{v} \in \mathbb{R}^l$ is an observed noisy image, $\mathbf{u} \in \mathbb{R}^l$ is the unknown clean image and $\mathbf{m} \in \mathbb{R}^l$ is white Gaussian noise. In the shrinkage-based approach, the restored image is given by the following algorithm

$$\hat{\mathbf{u}} = W^\dagger \overrightarrow{\Psi} (W\mathbf{v}), \tag{3}$$

where W is a unitary or overcomplete transform, $\overrightarrow{\Psi} = [\Psi_1, \Psi_2, \ldots]$ is a set of scalar shrinkage functions and $W^\dagger$ is the reverse transform. The utilization of scalar shrinkage functions is derived in the following subsections.

2.1   The Unitary Case

The shrinkage-based reconstruction (3) can be shown to solve a MAP estimation problem under the assumptions of a unitary transform, independent transform coefficients and white Gaussian noise. The discussion is focused on the unitary wavelet transform, since it provides a sparse representation of natural images [11] and its coefficients are assumed independent. These properties of the unitary wavelet transform play a fundamental role in the formulation of sparsity-promoting image priors [12] that can be decoupled into a product (or a sum in the log domain) of scalar probability distributions. The MAP estimator $\hat{\mathbf{u}}(\mathbf{v})$ is given by maximizing the a-posteriori probability:

$$\hat{\mathbf{u}}(\mathbf{v}) = \arg \max_{\mathbf{u}} P(\mathbf{u} \mid \mathbf{v}). \tag{4}$$

---

[1] In this paper we refer to the work in [1], [2] as "sparse representation" based, although the shrinkage based approach also emerges from sparse representation modeling.

This maximization can be cast also in the transform domain, as follows:

$$\hat{\mathbf{u}}_W\left(\mathbf{v}_W\right) = \arg\max_{\mathbf{u}_W} P\left(\mathbf{u}_W \mid \mathbf{v}_W\right), \tag{5}$$

where $\mathbf{u}_W = W\mathbf{u}$, $\mathbf{v}_W = W\mathbf{v}$ and W is a unitary wavelet transform. By utilizing *Bayes* rule and the monotonicity of the log function, the wavelet domain MAP estimator can be reformulated as

$$\hat{\mathbf{u}}_W\left(\mathbf{v}_W\right) = \arg\min_{\mathbf{u}_W}\{-\log P\left(\mathbf{v}_W \mid \mathbf{u}_W\right) - \log P\left(\mathbf{u}_W\right)\}. \tag{6}$$

The term $\log P\left(\mathbf{v}_W \mid \mathbf{u}_W\right)$ is the *log likelihood* and the term $\log P\left(\mathbf{u}_W\right)$ is the *prior*. For the white Gaussian noise case, the log likelihood term is given by

$$-\log P\left(\mathbf{v}_W \mid \mathbf{u}_W\right) = \lambda\|\mathbf{u}_W - \mathbf{v}_W\|^2 = \lambda\sum_i \|u_W^i - v_W^i\|^2, \tag{7}$$

where u_W^i and v_W^i are the i-th elements of $\mathbf{u}_W$ and $\mathbf{v}_W$, respectively and λ is a constant inversely proportional to the noise variance. Note, that by utilizing the l^2-norm preserving property of unitary transforms, equation (7) can be rewritten as

$$-\log P\left(\mathbf{v}_W \mid \mathbf{u}_W\right) = \lambda\|W\left(\mathbf{u} - \mathbf{v}\right)\|^2 = \lambda\|\mathbf{u} - \mathbf{v}\|^2 = -\log P\left(\mathbf{v} \mid \mathbf{u}\right). \tag{8}$$

Thus, the spatial domain MAP estimator (4) and its unitary transform domain counterpart (5) are equivalent, as long as the *prior* term is a function of $W\mathbf{u}$ [12]. By utilizing the independence assumption of the unitary wavelet coefficients, the *prior* term is reformulated as

$$\log P\left(\mathbf{u}_W\right) = \log \prod_i P_i\left(u_W^i\right) = \sum_i \log P_i\left(u_W^i\right). \tag{9}$$

The unitary wavelet domain MAP estimator (6) can be rewritten using the results of equations (7) and (9), leading to a decoupling of the l-dimensional minimization problem to a set of l scalar minimization problems

$$\hat{u}_W^i\left(v_W^i\right) = \arg\min_{u_W^i}\{\lambda\|u_W^i - v_W^i\|^2 - \log P_i\left(u_W^i\right)\} \quad \forall i. \tag{10}$$

The optimization in equation (10) is solved by applying a scalar lookup table function Ψ_W^i, termed *shrinkage* function, to the wavelet coefficients: $\hat{u}_W^i\left(v_W^i\right) = \Psi_W^i\left(v_W^i\right)$. The shrinkage function depends solely on the noise variance and the prior term $P_i\left(u_W^i\right)$. The pioneering studies of Donoho and Johnstone [13], [14] suggested using *hard-thresholding* and *soft-thresholding* shrinkage functions. Furthermore, for a K subband wavelet transform, only K distinct shrinkage functions are required to solve the MAP estimation problem. To clarify this property we follow the notation in [9] and utilize a permutation matrix P to reorder the rows of the wavelet transform W. The reordering is performed such that wavelet

transform rows corresponding to a specific subband are co-located in a distinct block

$$B = PW = \begin{bmatrix} B_1 \\ \vdots \\ B_K \end{bmatrix} \quad \text{and} \quad \mathbf{v_B} = B\mathbf{v} = \begin{bmatrix} \mathbf{v_{B_1}} \\ \vdots \\ \mathbf{v_{B_K}} \end{bmatrix}. \tag{11}$$

The set of K shrinkage functions are denoted by $\overrightarrow{\Psi_B} = [\Psi_{B_1}, \Psi_{B_2}, \ldots, \Psi_{B_K}]$ and the restored image (3) is given by

$$\hat{\mathbf{u}} = B^T \overrightarrow{\Psi_B}\{\mathbf{v_B}\} = \sum_{k=1}^{K} B_k^T \Psi_{B_k}\{\mathbf{v_{B_k}}\}, \tag{12}$$

where B^T is the reverse transform due to the unitary case assumption.

2.2   The Overcomplete Case

The shrinkage restoration approach in the unitary case provides good results, however, visual artifacts sometimes appear in the restored image. By utilizing an overcomplete transform, significant improvements can be achieved. This was originally discovered by Coifman and Donoho [15] where an undecimated wavelet transform provided superior shrinkage denoising results compared to the unitary case. This improvement was later demonstrated in various overcomplete transforms such a Curvelets [16], Contourlets [17], undecimated windowed DCT [9] and others. By applying equation (11) to the overcomplete case, the noisy image transform is given by $\mathbf{v_B} = B\mathbf{v}$. The overcomplete transform B is an $M \times l$ matrix where $M > l$. By modifying $\mathbf{v_B}$ using a vector of shrinkage functions $\overrightarrow{\Psi_B}\{\mathbf{v_B}\}$ it is desired that all the post-shrinkage overcomplete components be equal to the overcomplete transform components of the original (unknown) image

$$B\mathbf{u} = \overrightarrow{\Psi_B}\{\mathbf{v_B}\}. \tag{13}$$

The estimated image is reconstructed using the pseudo-inverse

$$\hat{\mathbf{u}} = \left(B^T B\right)^{-1} B^T \overrightarrow{\Psi_B}\{\mathbf{v_B}\} = \left(B^T B\right)^{-1} \sum_{k=1}^{K} B_k^T \Psi_{B_k}\{\mathbf{v_{B_k}}\}. \tag{14}$$

A key difference between the unitary and overcomplete cases is statistical dependence of the transform coefficients: the scalar shrinkage approach emerged from the independence assumption of the unitary wavelet coefficients, however, this assumption no longer holds in the overcomplete case. Traditionally, the unitary case shrinkage functions were applied also to the overcomplete case, however, the interband dependencies of the overcomplete transform coefficients should be taken into account. The most accurate approach to handle this issue is to design a set of multi-dimensional shrinkage functions, however, such approach is highly complex. The approach suggested in [9] for image denoising is to learn a set of *scalar* shrinkage functions that would take into account *interband* as well

as *intraband* dependencies. In this approach, the shrinkage functions are learned offline from an example set of pairs of clean and noisy images. In this paper, we leverage the shrinkage learning technique to the SISR problem and propose to learn the shrinkage functions online - from the observed image - in a way that would capture the statistical properties of (only) the image to be scaled-up.

2.3   From Image Denoising to Super-Resolution

The shrinkage-based restoration framework was originally developed for image denoising. However, it has been successfully utilized for more complex inverse problems then (2), by designing the shrinkage operation to minimize all structured noise components inherent to the specific problem. For example, inpainting by hard-thresholding [18], SISR by hard-thresholding [7], [8] and JPEG deblocking [9]. We Assume a general image degradation model:

$$\mathbf{v} = \Omega\{\mathbf{u}\} = \mathbf{u} + \mathbf{e}, \tag{15}$$

where $\Omega\{\cdot\}$ is a degradation operator (not necessarily linear) and $\mathbf{e}$ is an error image with unknown statistical properties. We propose to recover the unknown image $\mathbf{u}$ by utilizing the restoration algorithm (14), with a set of shrinkage functions that were designed to maximize the restored image quality, given the degradation model (15). For the SISR problem, we utilize the following degradation operator

$$\Omega\{\mathbf{u}\} = \Upsilon_\uparrow \left(DH\mathbf{u}\right), \tag{16}$$

where $\Upsilon_\uparrow \left(\cdot\right)$ is a simple interpolator (implemented either by a bi-linear or bi-cubic interpolator). Note, that this degradation operator simply amounts to an interpolation of the observed image $\mathbf{y}$ in the SISR model (1) and the dimensions of the degraded image $\mathbf{v} = \Omega\{\mathbf{u}\} = \Upsilon_\uparrow \left(DH\mathbf{u}\right)$ are identical to $\mathbf{u}$. Therefore, the proposed restoration scheme for the SISR problem is as follows

$$\hat{\mathbf{u}} = \left(B^T B\right)^{-1} B^T \overrightarrow{\Psi_B}\{\mathbf{v_B}\} = \left(B^T B\right)^{-1} \sum_{k=1}^{K} B_k^T \Psi_{B_k}\{\mathbf{v_{B_k}}\}. \tag{17}$$

In the proposed approach, the shapes of all shrinkage functions $\overrightarrow{\Psi_B}$ are trained for the SISR problem. The training is performed online (i.e. directly) from the observed image $DH\mathbf{u}$, exploiting the scale-invariant property of natural images [10]. The learning procedure relies on a piece-wise linear model of the shrinkage functions as explained in the following section. The learning process is explained in section 4.

3   The Slice Transform

The Slice Transform (SLT) [9] enables the approximation of a shrinkage function in a linear manner

$$\Psi_{B_k}\{\mathbf{v}_{B_k}\} \approx S_{\mathbf{q}_k}\left(\mathbf{v}_{B_k}\right)\mathbf{p}_k. \tag{18}$$

Note, that while the shrinkage function is a scalar function, the representation (18) incorporates the element-wise shrinkage operation for the entire subband B_k. The i-th row of the sparse matrix $S_{\mathbf{q}_k}(\mathbf{v}_{B_k})$ is determined uniquely by the i-th element of the vector $\mathbf{v}_{\mathbf{B_k}}$ and the predefined vector $\mathbf{q}_k$. The vector $\mathbf{p}_k$ is the design parameter that controls the input-output mapping relation of the k-th shrinkage function. In the following we explain the concept behind the representation (18) and begin, for simlicity, with the scalar case.

Assume $x \in [a, b)$ is a real value and the half open interval $[a, b)$ is divided into M slots. The boundaries of the slots are contained in the vector $\mathbf{q} = [q_0, q_1, \ldots, q_M]^T$ such that $q_0 = a < q_1 < q_2 \cdots < q_M = b$. The value x is located in a single slot $\pi(x) \in \{1, \ldots, M\}$ and associated with a residue $r(x)$, where $\pi(x) = j$ if $x \in [q_{j-1}, q_j)$ and

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

Note that $r(x) \in [0, 1)$, where $r(x) = 0$ if $x = q_{\pi(x)-1}$ and $r(x) \to 1$ if $x \to q_{\pi(x)}$. The value x can be expressed as a follows

$$x = S_{\mathbf{q}}(x)\mathbf{q} = r(x)q_{\pi(x)} + (1 - r(x))q_{\pi(x)-1}, \tag{19}$$

where the row vector $S_{\mathbf{q}}(x) \in \mathbb{R}^{(M+1)}$ is defined as follows:

$$S_{\mathbf{q}}(x) = [0, \ldots, 0, 1 - r(x), r(x), 0, \ldots, 0]$$

and where the values $1 - r(x)$ and $r(x)$ are located in the $(\pi(x) - 1)^{th}$ and $(\pi(x))^{th}$ entries, respectively. Extending equation (19) to the multi-dimensional case, we assume that $\mathbf{x} \in \mathbb{R}^N$ and that each element satisfies $x^i \in [a, b)$. The SLT of $\mathbf{x}$ is given by

$$\mathbf{x} = S_{\mathbf{q}}(\mathbf{x})\mathbf{q}, \tag{20}$$

where the matrix $S_{\mathbf{q}}(\mathbf{x}) \in \mathbb{R}^{N \times (M+1)}$ is given by

$$[S_{\mathbf{q}}(\mathbf{x})]_{i,j} = \begin{cases} r(x^i) & \text{if } \pi(x^i) = j \\ 1 - r(x^i) & \text{if } \pi(x^i) = j + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

Each row of the matrix $S_{\mathbf{q}}(\mathbf{x})$ is associated with a single element of the vector $\mathbf{x}$ and the representation (20) is composed of linear splines basis functions. According to [9], substituting the boundary vector $\mathbf{q}$ with a different vector $\mathbf{p}$ performs a piece-wise linear mapping of the values in $\mathbf{x}$

$$\mathbf{M}_{\mathbf{q},\mathbf{p}}(\mathbf{x}) = S_{\mathbf{q}}(\mathbf{x})\mathbf{p}, \tag{22}$$

where $\mathbf{M}_{\mathbf{q},\mathbf{p}}(\mathbf{x})$ performs linear mapping of the values $\{x^i \in [q_{j-1}, q_j)\}$ to the interval $[p_{j-1}, p_j)$, as depicted in Fig. 1. The substitution property (22) is the key principal behind the linear representation of the shrinkage functions (18).
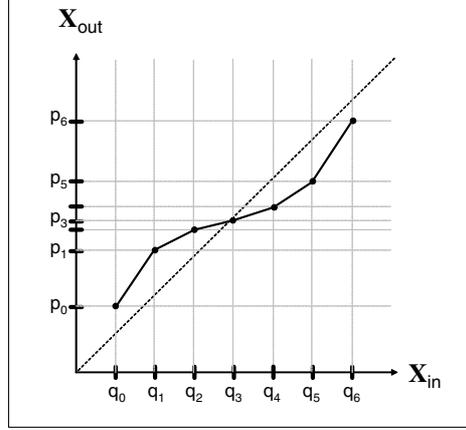
Fig. 1. Piece-wise linear mapping with the Slice Transform

4   The Super-Resolution Algorithm

The proposed Super-Resolution algorithm includes two stages: during the first stage a pair of example images are utilized in an online discriminative learning process of the shrinkage functions. In the second stage, the learned shrinkage functions are applied during the super-resolution reconstruction.

4.1   Stage I: Learning the Shrinkage Functions

The shrinkage functions learning algorithm is inspired by an *oracle* based approach. Consider the SISR degradation model (1) and an oracle estimator of the shrinkage functions that has access to the input image $\mathbf{y}$ and to the unknown high-resolution image $\mathbf{x}$. The oracle learning strategy is based on constructing a super-resolution approximation $\hat{\mathbf{x}}$ from the interpolated low-resolution image $\mathbf{y}_\uparrow = \Upsilon_\uparrow(\mathbf{y})$ by employing the scheme in (17) such that the unknown shrinkage functions are represented by the SLT approximation (18)

$$\hat{\mathbf{x}}(\mathbf{y}_\uparrow, \mathbf{p}) = \left(B^T B\right)^{-1} \sum_{k=1}^{K} B_k^T S_{\mathbf{q}_k}(\mathbf{y}_{\uparrow B_k}) \mathbf{p}_k = L(\mathbf{y}_\uparrow) \mathbf{p}, \tag{23}$$

where $\mathbf{p} = \left[\mathbf{p_1^T}, \mathbf{p_2^T}, \dots, \mathbf{p_K^T}\right]^{\mathbf{T}}$ and

$$\mathbf{L}(\mathbf{y}_\uparrow) = \left(\mathbf{B^T B}\right)^{-1} [\mathbf{H_1}, \mathbf{H_2}, \dots, \mathbf{H_K}]$$

where $\mathbf{H_i} = \mathbf{B_i^T S_{q_i}}(\mathbf{y_{\uparrow B_i}})$. The oracle learns the unknown shrinkage functions by solving the following Least Squares (LS) optimization problem

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}_\uparrow, \mathbf{p})\|_\mathbf{2}^\mathbf{2}. \tag{24}$$

This ideal strategy captures all interband and intraband statistical dependencies of the reconstructed image such that the spatial domain mean squared error (MSE) between the reconstructed and true images is minimized. In practice, only the observed low-resolution image is available and a question arises - can we learn the shrinkage functions in a similar fashion to the oracle with only $\mathbf{y}$ at hand? Here we exploit the scale-invariant property of natural images [10] and we approximate it by the following approach: the oracle training pair $\{\mathbf{x}, \mathbf{y}_\uparrow\}$ is replaced with the pair $\{\mathbf{y}, \mathbf{g}\}$ such that the reference image is now the observed low-resolution image and its degraded counterpart is given by

$$\mathbf{g} = \Upsilon_\uparrow \left(\widetilde{D}\widetilde{H}\mathbf{y} \right) \in \mathbb{R}^n,$$

where $\widetilde{\mathbf{H}} \in \mathbb{R}^{\mathbf{n} \times \mathbf{n}}$ is a blurring filter (block-circulant) convolution matrix and $\widetilde{\mathbf{D}} \in \mathbb{R}^{\frac{\mathbf{n}}{\mathbf{L}} \times \mathbf{n}}$ is a down-sampling operator matrix, by a factor of $\sqrt{L}$ along the horizontal and vertical dimensions. Thus, the super-resolution reconstruction is applied to a scaled-down version of the low-resolution observed image

$$\hat{\mathbf{y}}(\mathbf{g}, \mathbf{p}) = \left(B^T B\right)^{-1} \sum_{k=1}^{K} B_k^T S_{\mathbf{q}_k} \left(\mathbf{g}_{B_k}\right) \mathbf{p}_k = L(\mathbf{g})\,\mathbf{p}. \tag{25}$$

The shrinkage functions are jointly learned by solving the following LS problem

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \|\mathbf{y} - \hat{\mathbf{y}}(\mathbf{g}, \mathbf{p})\|_{\mathbf{2}}^{\mathbf{2}} \tag{26}$$

and the solution is given by

$$\hat{\mathbf{p}} = \left(\mathbf{L^T L}\right)^{-1} \mathbf{L^T y}, \tag{27}$$

where $\mathbf{L} = \mathbf{L}(\mathbf{g})$.

4.2   Stage II: Super-Resolution Reconstruction

Once the parameters of the shrinkage functions are learned, the super-resolution image is reconstructed as follows

$$\hat{\mathbf{x}}(\mathbf{y}_\uparrow, \hat{\mathbf{p}}) = \left(B^T B\right)^{-1} \sum_{k=1}^{K} B_k^T S_{\mathbf{q}_k} \left(\mathbf{y}_{\uparrow_{B_k}}\right) \hat{\mathbf{p}}_k. \tag{28}$$

5   Performance Evaluation

The performance of the proposed algorithm was evaluated by computer simulations and compared versus bi-cubic interpolation and the state-of-the-art sparse-representation based algorithm [1], [2] (which outperforms the sparse-representation approach [3]). Performance were not compared to the method [5]

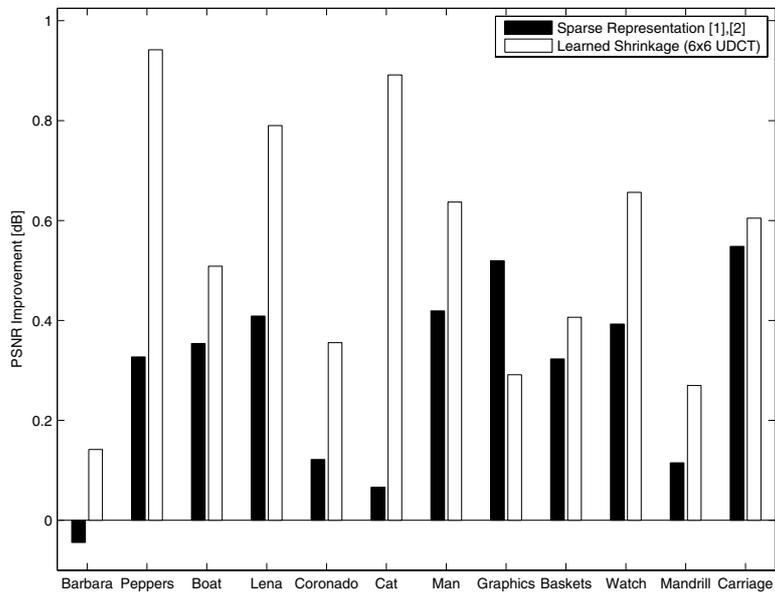**Fig. 2.** The collection of tested images, all images are of size 512×512

Fig. 3. PSNR improvement over bi-cubic interpolation for $L = 9$

Table 1. PSNR results for $L = 9$

Image	Bicubic Interp.	Sparse Rep. [1]	Shrink. 4×4	Shrink. 6×6	Shrink. 8×8	Offline Shrink.
Barbara	24.05	24.00	24.18	**24.19**	24.17	24.06
Peppers	29.82	30.14	30.51	**30.76**	30.75	29.77
Boat	27.14	27.49	27.58	**27.65**	27.61	27.31
Lena	30.76	31.17	31.45	**31.55**	31.53	30.85
Coronado	25.24	25.36	25.53	**25.59**	25.46	25.36
Cat	28.67	28.73	29.32	29.56	**29.59**	28.79
Man	28.35	28.77	28.92	**28.98**	28.91	28.46
Graphics	21.79	22.31	**22.36**	22.08	22.29	21.94
Baskets	21.13	21.45	21.48	**21.54**	21.51	21.20
Watch	28.10	28.49	28.60	**28.75**	28.68	28.18
Mandrill	22.01	22.13	22.24	22.28	22.27	**22.86**
Carriage	27.41	27.96	27.90	**28.02**	27.98	27.47
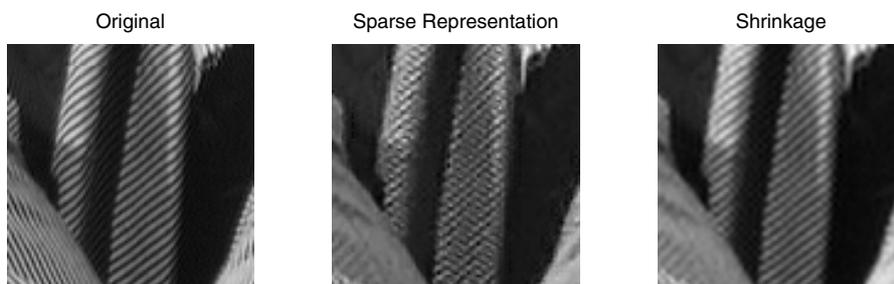
Original   Sparse Representation   Shrinkage

Fig. 4. Super-resolution of the image *barbara* for $L = 4$

Original   Sparse Representation   Shrinkage

Fig. 5. Super-resolution of the image *barbara* for $L = 4$

Original   Sparse Representation   Shrinkage

Fig. 6. Super-resolution of the image *watch* for $L = 4$

Original   Sparse Representation   Shrinkage

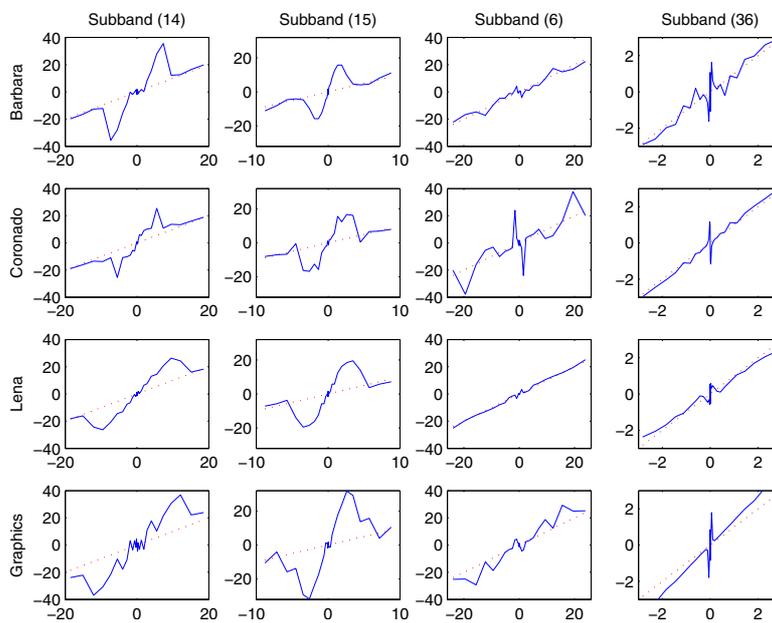**Fig. 7.** Super-resolution of the image *watch* for $L = 4$

Fig. 8. Examples of learned shrinkage functions for $L = 9$. In each row the image is fixed and in each column the subband is fixed.

as neither quantitative results nor a code of this method were available for evaluation. A collection of 12 images presented in Fig. 2 were compared against their own SISR reconstructions (from their scaled-down versions) with scale-up factors of $L = 4$ and $L = 9$. The undecimated $\sqrt{K} \times \sqrt{K}$ windowed DCT (UDCT) was chosen as the overcomplete transform. This transform is defined to include all possible $\sqrt{K} \times \sqrt{K} = K$ DCT window shifts, leading to a redundancy factor of K with K distinct subbands. In this approach, each subband $\mathbf{y}_{\uparrow_{B_k}} = B_k \mathbf{y}_{\uparrow}$ is generated by filtering the image with the respective basis kernel. In addition, the UDCT is a *tight frame* thus the term $\left(B^T B\right)^{-1}$ boils down to the identity matrix. PSNR results are compared in Fig. 3 for $L = 9$ and a 6×6 UDCT, where it can be seen that the proposed approach outperforms the sparse-representation approach for all the images (excluding the image *graphics*). The proposed approach achieved an average gain of 0.54dB over bi-cubic interpolation, versus an average gain of 0.30dB achieved by the sparse-representation approach. Detailed PSNR results are presented in Table 1 for all methods. Three different UDCT window sizes were compared for online learning and it can be seen that the 6×6 window size provided the best results. In addition, the offline learning approach [9] was evaluated by training the shrinkage functions with an image from the training collection reported in [9]. The offline training was performed using equation (24), with a 6×6 UDCT. It can be seen that the offline approach provided inferior results compared to the online approach (excluding the image *mandrill*). In the specific case of the image *graphics*, it is possible that the assumption of scale-invariance is not as true, explaining the lower performance obtained. These type of images could be treated using the offline approach with adequately chosen training examples. Visual comparison of SISR reconstructions are presented for $L = 4$ in Figs. $4-7$, it can be seen that various artifacts appear in the sparse-representation based approach while the proposed approach produces more natural and pleasant results (figures are best viewed in the electronic version of this paper). Examples of learned shrinkage functions are presented in Fig. 8, where it can be seen that for a fixed subband the learned shrinkage functions exhibit significantly different behavior for different images. For instance, only in subband (6) of the image *coronado* there is significant boosting effect with sign inversion for low amplitude coefficients.

6   Conclusions

This paper presented a novel approach for shrinkage functions learning in single image super-resolution. By exploiting the scale-invariant property of natural images, the set of scalar shrinkage functions are jointly learned from the low-resolution input image. Computer simulations with a simple overcomplete dictionary - the undecimated windowed DCT - revealed superior performance versus the state-of-the-art sparse-representation approach. Future research directions include a joint online-offline learning approach that combines additional example images into the online learning process. In addition we will consider the reconstruction of the residual error image in (15) rather then the complete image - thus focusing the learning process only into the missing high-pass components.

Acknowledgement

References

1. Yang, J., Wright, J., Ma, Y., Huang, T.: Image super-resolution as sparse representation of raw image patches. In: CVPR (2008)
2. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. Submitted to IEEE Trans. on Image Processing (2010)
3. Wang, J., Zhua, S., Gonga, Y.: Resolution enhancement based on learning the sparse association of image patches. Pattern Recognition Letters 31, 1–10 (2010)
4. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. IEEE Computer Graphics and Applications 22, 56–65 (2002)
5. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: ICCV (2009)
6. Elad, M., Datsenko, D.: Example-based regularization deployed to super-resolution reconstruction of a single image. The Computer Journal 50, 1–16 (2007)
7. Guleryuz, O.G.: Predicting wavelet coefficients over edges using estimates based on nonlinear approximants. In: Proceedings of the Data Compression Conference (2004)
8. Mueller, N., Lu, Y., Do, M.N.: Image interpolation using multiscale geometric representations. In: SPIE Symposium on Electronic Imaging, San Jose (2007)
9. Hel-Or, Y., Shaked, D.: A discriminative approach for wavelet denoising. IEEE Trans. on Image Processing 17, 443–457 (2008)
10. Ruderman, D.L., Bialek, W.: Statistics of natural images: Scaling in the woods. Phys. Rev. Lett. 73, 814–817 (1994)
11. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609 (1996)
12. Elad, M.: Why simple shrinkage is still relevant for redundant representations? IEEE Trans. on Information Theory 52, 5559–5569 (2006)
13. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455 (1994)
14. Donoho, D.L.: Denoising by soft thresholding. IEEE Trans. on Information Theory 41, 613–627 (1995)
15. Coifman, R.R., Donoho, D.L.: Translation invariant de-noising. In: Lecture Notes in Statistics: Wavelets and Statistics, pp. 125–150 (1995)
16. Ma, J., Plonka, G.: The curvelet transform. IEEE Signal Processing Magazine 27, 118–133 (2010)
17. Do, M.N., Vetterli, M.: The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans. on Image Processing 14, 2091–2106 (2005)
18. Guleryuz, O.G.: Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part i: theory. IEEE Trans. on Image Processing 15, 539–554 (2006)