1

Spatially adaptive filtering as regularization in inverse imaging: compressive sensing, super-resolution and upsampling

Aram Danielyan

Department of Signal Processing, Tampere University of Technology

Alessandro Foi

Department of Signal Processing, Tampere University of Technology

Vladimir Katkovnik

Department of Signal Processing, Tampere University of Technology

Karen Egiazarian

Department of Signal Processing, Tampere University of Technology

CONTENTS

The recent developments in image and video denoising have brought a new generation of filtering algorithms achieving unprecedented restoration quality. This quality mainly follows from exploiting various features of natural images. The nonlocal self-similarity and sparsity of representations are key elements of the novel filtering algorithms, with the best performance achieved by adaptively aggregating multiple redundant and sparse estimates. In a very broad sense, the filters are now able, given a perturbed image, to identify its plausible representative in the space or manifold of possible solutions. Thus, they are powerful tools not only for noise removal, but also for providing accurate adaptive regularization to many ill-conditioned inverse imaging problems. In the case of image/video reconstruction from incomplete data, the general structure of the proposed approach is very simple and is based on iteratively refining the estimates alternating two procedures: image/video filtering (denoising) and projection on the observation-constrained subspace. In this chapter we give an overview of this versatile approach, with particular emphasis on three challenging and important imaging applications: inversion from sparse or limited-angle tomographic projections, image reconstruction from low-frequency or undersampled data, image and video super-resolution. This approach is especially appealing for the latter application, as the block-matching procedure, performed both in space and time, makes the explicit motion estimation unnecessary. The presented experimental results demonstrate an overall performance on the level of the state of the art.

1.1   Introduction

A-priori assumptions on the image to be reconstructed are essential for any inverse imaging algorithm. In the standard variational approaches, these assumptions are usually given as penalty terms which serve as regularization in an energy criterion to be minimized.

A main limitation of these approaches is that the minimization usually involves the evaluation of the gradient of the penalty functional and its convexity. Therefore, if on the one hand the nonstationarity of natural images calls for locally adaptive non-convex penalties, on the other hand, to obtain a feasible algorithm, the energy criterion and, thus, the penalty need to be simple.

This limitation becomes evident particularly for image denoising, for which the recent years have witnessed the development of a number of spatially adaptive algorithms that dramatically outperform the established methods based on variational constraints (see, e.g., [20],[26]).

Our approach to inverse imaging essentially differs from these variational formulations and appeals to nonparametric regression techniques. We propose to replace the implicit regularization coming from the penalty by explicit

filtering, exploiting spatially adaptive filters sensitive to image features and details. If these filters are properly designed, we have reasonable hopes to achieve better results by filtering than through the formal approach based on the formulation of imaging as a variational problem with imposed global constraints.

The presented framework is general and is applicable to a wide class of inverse problems for which the available data can be interpreted as a smaller portion of some transform spectrum of the signal of interest. This observation model has gained recently enormous popularity under the paradigm of compressive sensing [1].

We demonstrate application of our approach to the inversion from sparse or limited-angle tomographic projections, image upsampling, and image/video super-resolution. Depending on the particular way of sampling the spectral components and the assumptions about complexity of the image, the recursive procedures can be improved by incorporating random search or staged reconstruction.

The chapter is organized as follows. In the next section we give formal definitions of the observation model and of the reconstruction algorithm for a rather general case. These are then reinterpreted in the context of compressive sensing in Section 1.3, where we also present inverse tomography and basic image upsampling experiments. Section 1.4 is devoted to image and video super-resolution. Concluding remarks are given in the last section.

1.2 Iterative filtering as regularization

Let us consider a general ill-posed inverse problem in the form

$$y = \mathcal{H}(x) \tag{1.1}$$

where x is the true image to be reconstructed and y is its observation through a linear operator $\mathcal{H} : X \rightarrow Y$, X and Y being Euclidean spaces with dimensions $\dim Y < \dim X$. We start from the naive pseudoinverse

$$\hat{x}^{(0)} = \underset{x:\mathcal{H}(x)=y}{\arg\min} \|x\|_2$$

of (1.1), which is usually very far from the solution one would like to obtain, particularly when $\dim Y$ is much smaller than $\dim X$. Nevertheless, due to the linearity of $\mathcal{H}$, any other solution of (1.1) should differ from $\hat{x}^{(0)}$ only by its component on the null space of $\mathcal{H}$, $\ker(\mathcal{H})$. To obtain an updated estimate $\hat{x}^{(1)}$, we first refine $\hat{x}^{(0)}$ with a filter Φ, project $\Phi(\hat{x}^{(0)})$ on $\ker(\mathcal{H})$ and then add $\hat{x}^{(0)}$. The reason of the last two operations is that the refined estimate, while being closer to the desired solution may not satisfy (1.1) anymore. This procedure

can be repeated iteratively leading to the following recursive scheme:

$$
\begin{cases}
\hat{x}^{(0)} = \underset{x:\mathcal{H}(x)=y}{\arg\min} \|x\|_2, \\
\hat{x}^{(k)} = \hat{x}^{(0)} + \mathcal{P}_{\ker(\mathcal{H})}\left(\Phi\left(\hat{x}^{(k-1)}\right)\right),
\end{cases}
\tag{1.2}
$$

where $\mathcal{P}_{\ker(\mathcal{H})}$ is the projection operator on the null space $\ker(\mathcal{H})$ and the superscripts denote the corresponding iteration.

1.2.1   Spectral decomposition of the operator

There are several ways how the projection $\mathcal{P}_{\ker(\mathcal{H})}$ can be realized. A practical approach, which we follow throughout the chapter, relies on the following spectral decomposition of the operator $\mathcal{H}$ [8].

Let $\mathcal{T} : X \to \Theta$ be an orthonormal transform with basis elements $\{t_i\}_{i=1}^{\dim X}$ such that $\ker(\mathcal{H}) = \mathrm{span}\,\{t_i\}_{i\in\Omega^C}$, Ω being a subset of indices (subband) and Ω^C its complementary. Given such a transform $\mathcal{T}$, the projection of $x \in X$, $\mathcal{P}_{\ker(\mathcal{H})}(x)$, can be easily obtained as the zeroing out of the $\mathcal{T}$-spectrum of x on Ω followed by application of $\mathcal{T}^{-1}$. Moreover, we can choose $\mathcal{T}$ so that it performs an eigendecomposition, i.e. so that $\mathcal{H}$ can be rewritten as

$$
\mathcal{H}(\cdot) = \mathcal{T}_0^{-1}\left(\mathcal{S}\left(\mathcal{T}(\cdot)\right)\right),
\tag{1.3}
$$

where $\mathcal{S} : \Theta \to \Theta_\Omega$ is a diagonal operator, which scales each spectrum coefficient by its corresponding non-zero eigenvalue, and $\mathcal{T}_0 : Y \to \Theta_\Omega$ is an orthonormal transform. Here Θ_Ω is the space obtained by restricting the elements of Θ on Ω. Thus, the operator $\mathcal{S}$ is simply restricting the $\mathcal{T}$-spectra on Ω and scaling the retained coefficients, with $\ker(\mathcal{S}) = \mathcal{T}(\ker(\mathcal{H}))$. Unless the eigenvalues are all distinct, the basis elements $\{t_i\}_{i\in\Omega}$ of the transform $\mathcal{T}$ that satisfies the above requirements are not uniquely determined. Of course, as the $\{t_i\}_{i\in\Omega}$ are varied, also the transform $\mathcal{T}_0$ must vary accordingly, because $\mathcal{T}_0$ is essentially determined by them. In matrix form, the spectral decomposition (1.3) can be obtained through the singular value decomposition (SVD) of $\mathcal{H}$**.**

1.2.2   Non-local transform domain filtering

In our implementations of the recursion (1.2), as the spatially adaptive filter Φ we utilize the Block-Matching 3D filtering (BM3D) denoising algorithm [10] and its extension to video V-BM3D [9]. Our choice of the BM3D algorithms is essentially determined by the fact that it is currently considered as one of the best denoising filters (see, e.g., [26],[20]) and that its remarkable performance is achieved at a competitive computational cost.

The BM3D algorithm exploits paradigms of non-local similarity [2],[20] in a blockwise fashion in order to obtain a highly sparse representation of the data. It means that the images to be processed are windowed/segmented into overlapping blocks and one looks for mutually similar blocks, which are
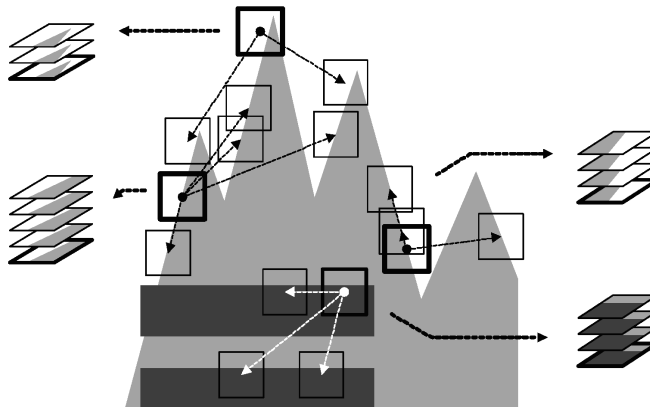
FIGURE 1.1
Illustration of grouping within an artificial image where for each reference block (with thick borders) there exist perfectly similar ones.

collected into groups, so that the data in these groups can be processed jointly. In this way we arrive to a non-local estimator with varying adaptive support where the data used in the estimation can be located quite far from each other. This estimation can be treated as a sophisticated high-order generalization of non-local means (NLM) [2],[4],[22],[23],[24].

To clarify the idea of grouping, let us consider an illustrative example of blockwise nonlocal estimation of the image in Figure 1.1 from an observation (not shown) corrupted by additive zero-mean independent noise. In particular, let us focus on the already grouped blocks shown in the same figure. These blocks exhibit perfect mutual similarity, which makes the elementwise averaging (i.e. averaging between pixels at the same relative positions) an optimal estimator. In this way, we achieve an accuracy that cannot be obtained by processing the separate blocks independently.

However, perfectly identical blocks are unlikely in natural images. If non-identical fragments are allowed within the same group, averaging is no longer optimal. Therefore, a filtering strategy more effective than averaging should be employed.

Here we give a brief description of the general V-BM3D algorithm, whose flowchart is presented in Figure 1.2. The BM3D algorithm can be then considered as equivalent to V-BM3D with the input sequence consisting of a single image. Detailed descriptions of the (V-)BM3D algorithms can be found in the corresponding references.

1. *Block-wise estimates.* Each image of an input sequence is processed in sliding-block manner. For each block the filter performs:

   (a) *Grouping by block-matching.* Searching within all images in the se-
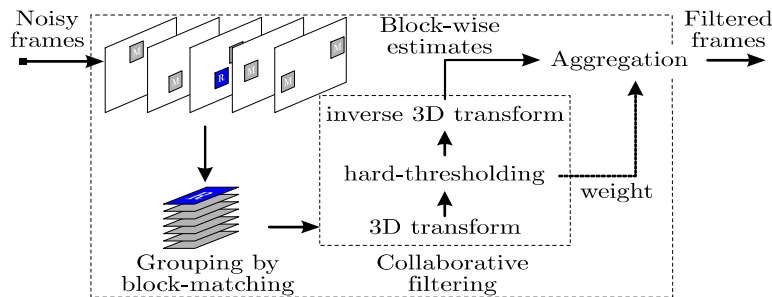
FIGURE 1.2
Flowchart of the the V-BM3D algorithm.

quence, find blocks that are similar to the currently processed one, and stack them together in a 3-D array (group).

(b) *Collaborative filtering.* Apply a 3-D transform to the formed group, attenuate the noise by hard-thresholding of the transform coefficients, invert the 3-D transform to produce estimates of all grouped blocks, and return the estimates of the blocks to their original place.

2. *Aggregation.* Compute the estimates of the output images by weighted averaging all of the obtained block-wise estimates that are overlapping.

What makes this algorithm very different from other non-local estimators is the use of a full-rank complete transform for modeling both the blocks and their mutual similarity or difference. Due to the similarity between the grouped blocks, the transform can typically achieve a highly sparse representation of the true signal so that the noise or small distortions can be well separated by shrinkage. In this way, the collaborative hard-thresholding reveals even the finest details shared by grouped fragments and at the same time it preserves the essential unique features of each individual fragment. Additionally, the adaptive aggregation of multiple redundant and sparse estimates allows to significantly boost the performance and provide high level of robustness to the algorithm.

Let us note that, for the purposes of this work, we do not perform the additional collaborative Wiener filtering stage which is otherwise present in the original BM3D and V-BM3D denoising algorithms [10], [9].

In what follows, the (V-)BM3D filter will be denoted as $\Phi(z, \sigma)$, where z is the degraded image/video to be filtered and $\sigma \geq 0$ is a parameter which originally corresponds to the standard-deviation of the noise in z and that here we mainly use to effectively control the filter strength: a larger σ implies a more aggressive filtering, thus enforcing a higher degree of sparsity, while a smaller σ provides better preservation of weak details at the expense of a milder noise suppression.

1.3   Compressed sensing

During the last three years, compressed sensing (CS) has received growing attention, mainly motivated by the positive theoretical and experimental results shown in [5], [7], [15], [19], [31], [32], [33]. The basic settings of signal reconstruction under conditions of CS are as follows. An unknown signal of interest is observed (sensed) through a limited number linear functionals. These observations can be considered as an incomplete portion of the spectrum of the signal with respect to a given linear transform $\mathcal{T}$. Thus, conventional linear reconstruction/synthesis (e.g., inverse transform) cannot in general reconstruct the signal. For example, when $\mathcal{T}$ is the Fourier transform, CS considers the case where the available spectrum is much smaller than what is required according to the Nyquist-Shannon sampling theory. It is generally assumed that the signal can be represented sparsely with respect to a different relevant basis (e.g., wavelets) or that, alternatively, it belongs to a specific class of functions (e.g., piecewise constant functions). Of particular importance is the so-called *incoherence* between the basis with respect to which the incomplete observations are given and the one with respect to which the signal is sparse [6]. In the publications cited above, it is shown that under such assumptions, stable reconstruction of the unknown signal is possible and that in some cases the reconstruction can be exact. These techniques typically rely on convex optimization with a penalty expressed by the ℓ_0 or ℓ_1 norm [34] which is exploited to enable the assumed sparsity [14]. It results in parametric modeling of the solution and in problems that are then solved by mathematical programming algorithms.

Based on the ideas discussed in the introduction, an alternative and novel approach to CS reconstruction can by developed by replacing the parametric modeling with a nonparametric one implemented by the use of spatially adaptive denoising filtering. The algorithm proposed in [17] extends the iterative scheme (1.2) by incorporating stochastic approximation to obtain stable recovery of the images. At every iteration, the current estimate is excited by injection of random noise in the unobserved portion of the spectrum. The denoising filter attenuates the noise and reveals new features and details out of the incomplete and degraded observations. Roughly speaking, we seek for the solution (reconstructed signal) by stochastic approximations whose search direction is driven by the denoising filter. It should be remarked that here we are concerned only in the operative way the solution is found, while we exploit essentially the same fundamental ideas of sparsity and basis incoherence undertaken by many other authors. As a matter of fact, the observation domains considered in what follows are much incoherent with respect to the adaptive data-driven 3-D transform domain in which BM3D effectively forces the data to be sparse.

1.3.1   Observation model and notation

Let $\mathcal{T} : X \rightarrow \Theta$ be an orthonormal transform operating from the image domain X to the transform domain Θ. The unknown image $x \in X$ can be sensed through the linear operator $\mathcal{H} : X \rightarrow Y = \Theta_\Omega$, where Ω is a subset of the $\mathcal{T}$-spectral components we are able to sense and Θ_Ω is the corresponding space of the $\mathcal{T}$-spectra restricted on Ω. The space Θ_Ω can be identified with the subspace of Θ constituted by all spectra that are identically zero outside of Ω. To clarify these concepts and simplify the coming notation, we introduce two operators:

- the *restriction operator* $|_\Omega$ that, from a given $\mathcal{T}$- spectrum, extracts its smaller portion defined on Ω;

- the *zero-padding operator* $\mathcal{U}_\Omega$ that expands the part of $\mathcal{T}$- spectrum defined on Ω to the full $\mathcal{T}$-spectrum by introducing zeros in the complement of Ω, Ω^C.

Likewise, we can define $|_{\Omega^C}$ and $\mathcal{U}_{\Omega^C}$.

Using these operators, the sensing operator can be explicitly written as $\mathcal{H}\left(\cdot\right) = \mathcal{T}\left(\cdot\right)|_\Omega$. Referring to (1.3), we have $\mathcal{S} = |_\Omega$ and $\mathcal{T}_0$ is the identity.

Now, if $x \in X$ is the unknown image intensity and $\theta = \mathcal{T}\{x\} \in \Theta$ is its spectrum, the CS problem is to reconstruct θ (or equivalently x) from the measurements y,

$$y = \theta|_\Omega = \mathcal{T}\left(x\right)|_\Omega . \tag{1.4}$$

It means that given the known part of the spectrum defined on Ω we have to reconstruct the missing part defined on Ω^C.

1.3.2   Iterative algorithm with stochastic approximation

Following (1.2), we obtain the initial estimate as $\hat{x}^{(0)} = \underset{x:\,\mathcal{H}(x)=y}{\arg\min} \|x\|_2 = \mathcal{T}^{-1}\left(\mathcal{U}_\Omega\left(y\right)\right)$. When the observed data is highly undersampled, the initial estimate $\hat{x}^{(0)}$ may contain too little information to enable reconstruction by the simple scheme (1.2). An improved scheme can be obtained by incorporating stochasticity into (1.2) resulting in

$$\begin{cases} \hat{x}^{(0)} = \underset{x:\,\mathcal{H}(x)=y}{\arg\min} \|x\|_2 = \mathcal{T}^{-1}\left(\mathcal{U}_\Omega\left(y\right)\right), \\ \hat{x}^{(k)} = \hat{x}^{(0)} + \mathcal{P}_{\ker(\mathcal{H})}\left[\left(1-\gamma_k\right)\hat{x}^{(k-1)} + \gamma_k\left(\Phi\left(\hat{x}^{(k-1)}, \sigma_k\right) + \eta_k\right)\right], k \geq 1. \end{cases} \tag{1.5}$$

Here, η_k is some pseudo-random white noise which is injected into the system in order provide the stochastic excitation. Equivalently, we can rewrite (1.5) with respect to spectral variables as

$$\begin{cases} \hat{\theta}^{(0)} = \underset{\theta:\,\theta|_\Omega=y}{\arg\min} \|\theta\|_2 = \mathcal{U}_\Omega\left(y\right), \\ \hat{\theta}^{(k)} = \hat{\theta}^{(0)} + \mathcal{P}_{\ker(\,|_\Omega)}\left[\left(1-\gamma_k\right)\hat{\theta}^{(k-1)} + \gamma_k\left(\mathcal{T}\left(\Phi\left(\mathcal{T}^{-1}\left(\hat{\theta}^{(k-1)}\right), \sigma_k\right)\right) + \eta_k\right)\right], k \geq 1, \end{cases} \tag{1.6}$$
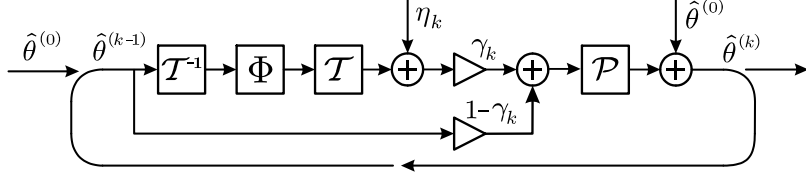
FIGURE 1.3
Flowchart of the recursive system (1.6).

where $\mathcal{P}_{\ker\left(\left.\right|_{\Omega}\right)}\left(\cdot\right) = \mathcal{U}_{\Omega^C}\left(\left(\cdot\right)|_{\Omega^C}\right)$ is the corresponding projection operator in transform domain which basically zeroes out all spectral coefficients defined on Ω.

The flowchart of the system (1.6) is shown in Figure 1.3. The system is initialized by setting $\hat{\theta}^{(0)} = \mathcal{U}_{\Omega}\left(y\right)$. Then, each iteration ($k \geq 1$) comprises of the following steps:

- *Image-domain estimate filtering.* The spectrum estimate is inverted and obtained image is filtered with the filter Φ. In this way, the coefficients from the unobserved part of the image spectrum are recreated from the given $\theta|_{\Omega} = y$. In our implementation of the algorithm the BM3D is used as the filter Φ.

- *Noise addition (excitation).* Pseudo-random noise η_k is introduced in the unobserved portion of the spectrum and works as a random generator of the missing spectral components. During the subsequent iterations, these components are attenuated or enhanced by the action of the filter Φ, depending to what extent they agree with the image features enabled by the observed spectrum $\theta|_{\Omega} = y$.

- *The updated* estimate $\hat{\theta}^{(k)}$ is obtained as the sum of the convex combination between the previous estimate $\hat{\theta}^{(k-1)}$ and the noise-excited prediction of the spectrum obtained after filtering, projected on the subspace defined by the unknown part of the spectra, and $\hat{\theta}^{(0)}$. The factor γ_k, in the convex combination, controls the rate of evolution (step size) of the algorithm.

1.3.2.1   Comments to the algorithm

Stochastic approximation. The iterative algorithm (1.6) can be treated as the Robbins-Monro stochastic approximation procedure (see, e.g., [25]). If the step-size parameter γ_k satisfies the standard conditions

$$\gamma_k > 0, \qquad \sum \gamma_k = \infty, \qquad \sum \gamma_k^2 < \infty \qquad (1.7)$$

and some assumptions on the operator $\mathcal{T}\left(\Phi\left(\mathcal{T}^{-1}\left(\cdot\right)\right)\right)$ hold, as $k \to \infty$ the estimates $\hat{\theta}^{(k)}$ from the recursive system (1.6) converge in mean squared sense to a solution $\hat{\theta}$ of the equation

$$E\left\{\mathcal{P}_{\ker\left(\mid_\Omega\right)}\left[-\hat{\theta}+\mathcal{T}\left(\Phi\left(\mathcal{T}^{-1}\left(\hat{\theta}\right),\sigma_k\right)\right)+\eta_k\right]\right\}=0,$$

i.e.

$$\mathcal{P}_{\ker\left(\mid_\Omega\right)}\left[\hat{\theta}-\mathcal{T}\left(\Phi\left(\mathcal{T}^{-1}\left(\hat{\theta}\right)\right)\right)\right]=0$$

or equivalently

$$\hat{\theta}=\mathcal{U}_\Omega\left(y\right)+\mathcal{P}_{\ker\left(\mid_\Omega\right)}\left[\mathcal{T}\left(\Phi\left(\mathcal{T}^{-1}\left(\hat{\theta}\right)\right)\right)\right]. \qquad (1.8)$$

If there is no smoothing in the filter Φ, the equation (1.8) becomes the identity. Thus, any $\hat{\theta}$ that satisfies observation (1.4) satisfies also to the equation (1.8), there is no image reconstruction and the algorithm does not work. Therefore, in order for the solution $\hat{\theta}$ to be non-trivial, the adaptive smoothing in (1.8) should be strong enough.

Excitation noise. The additive noise η_k used in the procedure (1.6) does not influence the equation (1.8). There are two arguments in favor of excitation of the algorithm by the random noise. First of all it improves the performance of the algorithm. It accelerates the transition process of the recursive procedure bringing it fast in the area of solution where the random walks steadies. The amplitude of these random walks decreases together with γ_k. It is well known (e.g., [21]) that the random search applied in optimization problems results in random walks well concentrated in areas of global extremum. Thus, the random search imposed by random excitation of the search trajectory can be useful for separation of local and global extrema. In a similar way, if the equation (1.8) has more than one solution, the randomness can help to find a "strong" solution with better quality of imaging or lower values of some hypothetical criterion function where the gradient (or quasi-gradient) can be defined as the vector corresponding to $\mathcal{P}_{\ker\left(\mid_\Omega\right)}\left(\hat{\theta}-\mathcal{T}\left(\Phi\left(\mathcal{T}^{-1}\left(\hat{\theta}\right)\right)\right)\right)$. Further, by changing the variance of the additive noise η_k one can control the rate of evolution of the algorithm. Thus, in practice, the assumptions (1.7) can be relaxed and a fixed γ_k can be used provided that $\mathrm{var}\left\{\eta_k\right\}\underset{k\to\infty}{\to}0$.

Filter parameters. The parameter σ_k is used in place of the standard-deviation of the noise of BM3D filter. This parameter controls the strength of collaborative hard-thresholding and thus affects the level of smoothing introduced by the filter. In order to prevent smearing of the small details the sequence $\left\{\sigma_k\right\}_{k=0,1,\ldots}$ should be decreasing with the progress of the iterations, and normally it is selected to be $\sigma_k^2=\mathrm{var}\left\{\eta_k\right\}$. The last fact makes (1.6) to formally differ from the classical Robbins-Monro procedure, where the operator Φ is assumed to be fixed (thus without a second argument σ) with the overall aggressiveness of the recursion controlled instead by the step size parameter γ_k.

Stopping rule. The algorithm can be stopped when the estimates $\hat{\theta}^{(k)}$ approach numerical convergence or after a specified number of iterations.

Image estimates. An image estimate $\hat{x}^{(k)}$ can be obtained at kth iteration as $T^{-1}\left(\hat{\theta}^{(k)}\right)$, although in practice $T^{-1}\left(\hat{\theta}^{(k)} - \mathcal{P}_{\ker\left(_{|\Omega}\right)}\left(\eta_{k-1}\right)\right)$ are better estimates because of the absence of excitation noise. All these estimates converge to $\hat{x} = T^{-1}\left(\hat{\theta}\right)$ as $k \to \infty$.

1.3.3   Experiments

The effectiveness of the proposed algorithm can be illustrated on two important inverse problems from computerized tomography: Radon inversion from sparse projections and limited-angle tomography. The former problem has been used as a benchmark for testing CS reconstruction algorithms (e.g., [5]). In particular, the results show that the presented algorithm allows to achieve exact reconstruction of synthetic phantom data even from a very limited number of projections. An example of image reconstruction from low-frequency data is also given. This particular example will serve as a bridge to the super-resolution problems considered in the next section.

The following experiments are carried out using a simplified form of the iterative scheme (1.6), where $\gamma_k \equiv 1$ and η_k is independent Gaussian noise with exponentially decreasing variance var $\{\eta_k\} = \alpha^{-k-\beta}$. For the filter Φ, we use the block-matching and 3-D filtering algorithm (BM3D) [10], setting $\sigma_k^2 = \text{var}\{\eta_k\}$. The separable 3-D Haar wavelet decomposition is adopted as the transform utilized internally by the BM3D algorithm.

We begin with illustrative inverse problems of compressed sensing for computerized tomography. As in [5], we simulate the Radon projections by "approximately" radial lines in the rectangular FFT domain. Note that the initial image estimate $\hat{x}^{(0)} = \underset{x:\mathcal{H}(x)=y}{\arg\min} \|x\|_2 = \mathcal{T}^{-1}\left(\underset{\theta:\theta|_\Omega=y}{\arg\min} \|\theta\|_2\right)$ coincides with the conventional back-projection estimate.
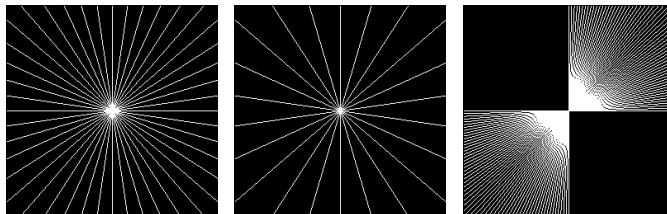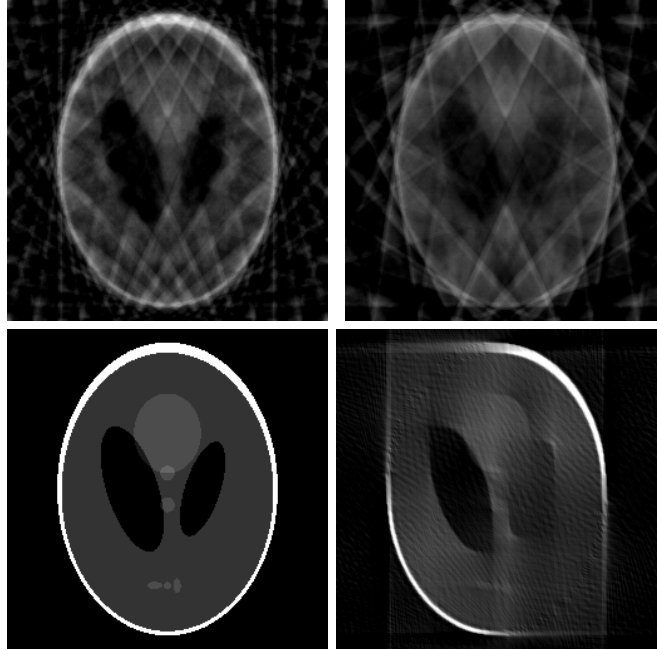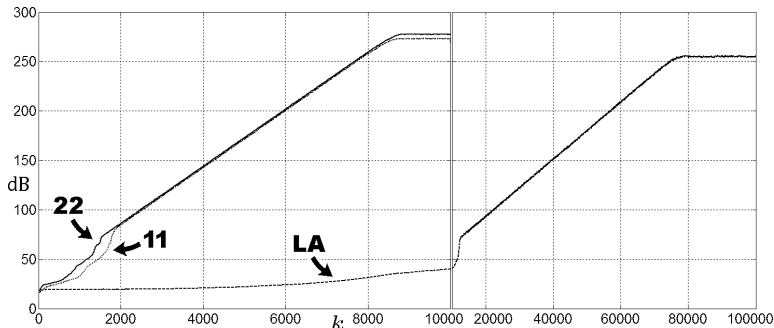

FIGURE 1.4
Available portion Ω of the FFT spectrum for the three experiments: 22 radial lines, 11 radial lines, 90 degrees limited-angle with 61 radial lines.

FIGURE 1.5
Clockwise from top-left: back-projection estimates for 22 radial lines, 11 radial lines, 61 radial lines with limited-angle (90 degrees), and original phantom (unknown and shown here only as a reference). For all three experiments, the estimates obtained after convergence of the algorithm coincide with the original image.

1.3.3.1   Radon inversion from sparse projections

First, we reproduce exactly the same experimental setup from [5], where 22 radial lines are sampled from the FFT spectrum of the Shepp-Logan phantom (size 256×256 pixels), as shown in Figure 1.4(left). Further, we reduce the number of available Radon projections from 22 to 11 (see Figure 1.4(center)). The initial back-projection estimates are shown in Figure 1.5. As the recursive algorithm progresses, the reconstruction error improves steadily until numerical convergence, as it can be seen from the plots in Figure 1.6. For both cases the reconstruction is *exact*, in the sense that the final reconstruction error (PSNR $\simeq$270dB) is comparable with the numerical precision of this particular implementation of the algorithm (double precision floating-point). We remark however that in practice such a high accuracy is never needed: already at a PSNR of about 60dB the image estimates can hardly be distinguished from the original.

FIGURE 1.6
Progression of the PSNR (dB) of the reconstructed image estimate $\hat{x}^{(k)}$ with respect to the iteration count k for the three experiments: 22 and 11 sparse projections ("22" and "11") and limited-angle ("LA").

1.3.3.2 Limited-angle tomography

In the two previous experiments, the available Radon projections were uniformly distributed with respect to the projection angle. A more difficult case arises when the angles under which the projections are taken are limited. Similarly to [27], we consider an overall aperture for the projections of 90 degrees. This restriction is essential, since all frequency information is completely missing along half of the orientations, which makes the reconstruction of, e.g., edges across these orientation extremely hard. We complicate the problem further, by taking only a smaller subset of 61 projections (a total of 256 properly-oriented projections would be required to cover a 90 degrees aperture). These sparse, limited-angle projections are illustrated in Figure 1.4(right). Although the convergence is here much slower than in the previous two experiments, the algorithm eventually achieves exact reconstruction.

In the above three experiments, as soon as the estimate reaches a quality of about 70dB, the recursion enters a phase of improvement at a constant rate (linear in terms of PSNR since $\mathrm{var}\{\eta_k\}$ decreases exponentially) which appears to be limited only by the used arithmetic precision.

1.3.3.3 Reconstruction from low-frequency data

The proposed recursive procedure can be applied also to more conventional image-processing problems. As a prelude to the image/video super-resolution and upsampling algorithms, which are the subject of the next section, we present here a basic experiment which can be seen as a prototype of image upsampling. In particular, we consider the reconstruction of a non-synthetic test image, namely *Cameraman* (256×256 pixels), from the low-frequency portion of its Fourier spectrum, with the set Ω being a 128×128 square centered at the DC, as illustrated in Figure 1.7. In Figure 1.8 we show the initial esti-
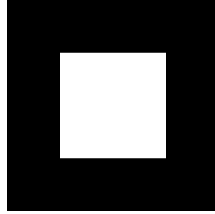
FIGURE 1.7
Available portion Ω of the FFT spectrum of the *Cameraman* image shown in Figure 1.8(left).

FIGURE 1.8
Cameraman: initial estimate $\hat{x}^{(0)}$ (zero-padding) (PSNR=27.32dB), and reconstructed estimate $\hat{x}^{(62)}$ after 62 iterations (PSNR=29.10dB).

mate $\hat{x}^{(0)}$ (by zero-padding in FFT domain, thus minimum ℓ_2-norm) and the reconstructed image obtained after few iterations of the algorithm. Despite the reconstruction is not exact, the salient details of the image are properly restored and there are no significant artifacts (e.g., ringing) thanks to the adaptivity embedded in the BM3D filter.

1.4 Super-resolution

Image *upsampling* or *zooming*, can be defined as the process of resampling a *single* low-resolution (LR) image on a high-resolution (HR) grid. Different resampling methods can be used to obtain zoomed images with specific desired

properties, such as edge preservation, degree of smoothness, etc. In particular, we concluded the previous section with an example demonstrating how one can achieve image upsampling by means of iterative BM3D filtering. As a matter of fact, because of the Nyquist-Shannon theorem, the inverse Fourier transform of the central low-frequency portion of the spectrum, can be treated as the LR image obtained upon decimation of the low-passed HR image.

However, fine details missing or distorted in the low-resolution image cannot be reconstructed in the upsampled one. This is particularly evident when the downsampling ratio between the HR and LR images is high (in the above example this ratio was only 2). Roughly speaking, there is no sufficient information in the low-resolution image to do this. The situation changes when a number of LR images portraying slightly different views of the same scene are available. The reconstruction algorithm now can try to improve the spatial resolution by incorporating into the final HR result the additional new details revealed in each LR image. The process of combining a sequence of undersampled and degraded low-resolution images in order to produce a single high-resolution image is commonly referred to as *super-resolution (SR) image reconstruction*, or, simply, *image super-resolution*.

The standard formulation of the image SR problem assumes that the LR images are obtained through the observations given by the model

$$x_{\mathrm{low}r} = \mathcal{D}\left(\mathcal{B}\left(\mathcal{F}_r\left(x_{\mathrm{hi}}\right)\right)\right), \quad r = 1, \ldots, R, \qquad (1.9)$$

where $x_{\mathrm{low}r}$ are LR images, $\mathcal{F}_r$, $\mathcal{B}$, $\mathcal{D}$ are the linear operators representing respectively warp, blur and decimaton, and x_{hi} is a high-resolution image of the scene subject to reconstruction. Blur and decimation operators are considered to be known.

Besides the difficulty involved in estimating the warping parameters, a principal drawback of the SR formulation (1.9) is that it assumes that all LR images $x_{\mathrm{low}r}$ can be mapped and upsampled to a unique HR image x_{hi}.

The observation model (1.9) can be then extended to the more general form

$$x_{\mathrm{low}r} = \mathcal{D}\left(\mathcal{B}\left(x_{\mathrm{hi}r}\right)\right), \quad r = 1, \ldots, R, \qquad (1.10)$$

where the warp operators on a single HR image are replaced by a sequence of HR images. This allows to consider arbitrary types of deformation between the frames, such as relative motion between different objects in the scene, and occlusions. The reconstruction of the HR sequence $\{x_{\mathrm{hi}r}\}_{r=1}^{R}$ from the LR sequence $\{x_{\mathrm{low}r}\}_{r=1}^{R}$ is termed *video super-resolution*.

The classical SR approach is loosely based on the following three steps: 1) registration of the LR images to a HR coordinate grid, 2) warping of the LR images onto that grid by interpolation, and 3) fusion of the warped images into the final HR image. An additional deblurring step is sometimes considered to compensate the blur. Several algorithms based on such classical approach exist and detailed reviews can be found, e.g., in [18].

For successful reconstruction it is crucial to perform accurate registration between the features represented across different frames. Most of the

existing SR methods rely either on a parametric global motion estimation, or on a computationally intensive optical flow calculation. However, an explicit registration of the LR frames is often not feasible: on the one hand, if the registration map has few degrees of freedom, it is too rigid to model the geometrical distortions caused by the lens system; on the other hand, when many degrees of freedoms are available (e.g., a dense optical flow), reliable estimates of the registration parameters cannot be obtained. In either case, registration artifacts are likely to appear in the fusion, requiring heavy regularization (smoothing) for their concealment [18]. The situation becomes even more difficult when non-global motion is present in images, something that is typical of video SR. Modern SR methods depart from the classical approach and we specially mention the video SR reconstruction algorithms [16], [29], [28], [30] based on the nonlocal means (NLM) filtering paradigm [2]. In these algorithms, instead of trying to obtain an explicit registration as a one-to-one pixel mapping between frames, a one-to-many mapping is utilized, where multiple pixels can be assigned to a given one, with weights typically defined by the similarity of the patches/blocks surrounding the pixels. The HR image is estimated through a weighted average of these multiple pixels (or of their surrounding patches) with their corresponding weights. The increased redundancy of the NLM, which can exploit also multiple patches from a same frame, contributes significantly to the overall good performance of the methods [29].

Here, based on our previous works [12],[13] we present a unified algorithm for the upsampling and image/video SR based on iterative (V)-BM3D filtering. As discussed in Section 1.2.2, the (V)-BM3D algorithm shares with the NLM the idea of exploiting nonlocal similarity between blocks. However, thanks to its transform-domain modeling, the BM3D turns out to be a much more effective filter than the NLM, thus leading to outstanding SR results.

The remainder of this section is organized as follows. First, we show how the SR problem can be interpreted in terms of the spectral decomposition (1.3). Then we present an adaptation of the algorithm (1.5) to video SR reconstruction, with image upsampling treated as a particular case. Finally, we report experimental results demonstrating the effectiveness and superior performance of the proposed approach.

1.4.1   Spectral decomposition for the super-resolution problem

The observation model (1.10) for the video SR problem is a particular case of the general observation model (1.1) where X is the space of high-resolution images, Y that of the low-resolution frames, and

$$\mathcal{H}\left(\cdot\right) = \mathcal{D}\left(\mathcal{B}\left(\cdot\right)\right). \tag{1.11}$$

Let us show how the blur and decimation operator relate to the general spectral decomposition presented in Section 1.2.1.

The blur in (1.10) can be written as the integral

$$\mathcal{B}\left(x_{\mathrm{hir}}\right)\left(s\right) = \int x_{\mathrm{hir}}\left(\xi\right) b\left(s, \xi - s\right) d\xi,$$

where b is a spatially varying point-spread function (PSF). Further, it can be represented as the inner products

$$\mathcal{B}\left(x_{\mathrm{hir}}\right)\left(s\right) = \langle x_{\mathrm{hir}}, b_s \rangle, \tag{1.12}$$

where $b_s\left(\xi\right) = b\left(s, \xi - s\right)$. Now, the action of the decimation operator $\mathcal{D}$ in (1.10) can be seen as a retaining only a subset of inner products (1.12) corresponding to the given sampling points s.

Let us assume that the corresponding retained b_s constitute a basis $\{t_i\}_{i=1}^{\dim Y}$ for a linear subspace $\tilde{X} \subset X$ of dimension equal to the number of pixels of any of the LR frames $x_{\mathrm{low}r}$. For the sake of simplicity, we assume this basis to be orthonormal up to a scaling constant β. Hence, $\tilde{X}$ is the orthogonal complement of $\ker\left(\mathcal{H}\right)$ in X.

The core of our modeling is to complete the basis $\{t_i\}_{i=1}^{\dim Y}$ with an orthonormal basis $\{t_i\}_{i=\dim Y+1}^{\dim X}$ of $\ker\left(\mathcal{H}\right)$, thus obtaining a basis $\{t_i\}_{i=1}^{\dim X}$ for X. This constitutes an orthonormal transform $\mathcal{T} : X \rightarrow \Theta$ whose basis elements separate $\ker\left(\mathcal{H}\right)$ from its orthogonal complement. By such a construction, the subset of basis indices $i = 1, \ldots, \dim Y$ correspond to the set Ω, $\Theta_\Omega = Y$. Hence, $\mathcal{T}_0$ is identity and $\mathcal{S}$ operates the division by β and restriction on Ω.

An evident example of this construction is when $\mathcal{B}$ is the blurring with a uniform kernel of size $n \times n$ and $\mathcal{D} = \downarrow_n$ is a decimation with rate n (along the rows as well as along the columns): the PSFs b_s, i.e. the basis elements $\{t_i\}_{i=1}^{\dim Y}$, are thus non-overlapping (hence orthogonal) shifted copies of the same uniform kernel. This basis can be naturally seen as the subbasis of the $n \times n$ block-DCT composed by extracting the DC-basis elements of all blocks. In this sense, we can complete $\{t_i\}_{i=1}^{\dim Y}$ with the basis $\{t_i\}_{i=\dim Y+1}^{\dim X}$ composed by the AC-basis elements of all blocks.

However, we are not bound to use the transform $\mathcal{T}$ constructed from the PSFs in the above direct way. As observed at the end of Section 1.2.1, the uniqueness of the basis elements $\{t_i\}_{i\in\Omega}$ depends on the non-zero eigenvalues being distinct, while here all non-zero eigenvalues equal β^{-1}. In this case, any orthonormal transform that provides the same orthogonal separation between $\ker\left(\mathcal{H}\right)$ and its complementary can be used in the decomposition. Indeed, in the example in Section 1.3.3.3 we have used the FFT as the transform $\mathcal{T}$ while the blur PSFs were sinc functions. Of course, if the used $\mathcal{T}$ differs from that constructed directly from the PSFs, also the transform $\mathcal{T}_0$ must differ from the identity. Therefore, in all the following equations, $\mathcal{T}_0$ is always written explicitly, and (1.11) takes thus the form

$$\mathcal{H}\left(\cdot\right) = \mathcal{T}_0^{-1}\left(\beta^{-1}\,\mathcal{T}\left(\cdot\right)|_\Omega\right). \tag{1.13}$$

Before we proceed further, let us remark that the assumption on the orthonormality (up to scaling constant) of the basis $\{t_i\}_{i=1}^{\dim Y}$ for $\tilde{X}$ induced by the blur and decimation operators is mainly for simplicity of exposition. If $\{t_i\}_{i=1}^{\dim Y}$ were not orthonormal, the construction would have been similar, differing from (1.13) either by having a frame and its dual instead of the orthonormal transform $\mathcal{T}$, or, in accordance with the general spectral decomposition (1.3), by substituting the factor β^{-1} by a general diagonal operator (i.e. multiplying each transform coefficient by its own scaling factor).

1.4.2 Observation model

Using the above representation of $\mathcal{H}$ (1.13), we reformulate the super-resolution observation model (1.10) as follows.

Let us be given a sequence of $R \geq 1$ *low-resolution images* $\{x_{\mathrm{low}r}\}_{r=1}^{R}$ of size $n_0^{\mathrm{h}} \times n_0^{\mathrm{v}}$, with each $x_{\mathrm{low}r}$ being obtained from the subband of the corresponding $\mathcal{T}$-spectra of original *higher-resolution* images $\{x_{\mathrm{hi}r}\}_{r=1}^{R}$ of size $n^{\mathrm{h}} \times n^{\mathrm{v}}$ as

$$y_r = x_{\mathrm{low}r} = \mathcal{T}_0^{-1}\left(\beta^{-1}\,\mathcal{T}\left(x_{\mathrm{hi}r}\right)|_{\Omega}\right). \tag{1.14}$$

The problem is to reconstruct $\{x_{\mathrm{hi}r}\}_{r=1}^{R}$ from $\{x_{\mathrm{low}r}\}_{r=1}^{R}$. Clearly, for a fixed r, any good estimate $\hat{x}_r$ of $x_{\mathrm{hi}r}$ must have its Ω subband equal to $\beta\mathcal{T}_0\left(x_{\mathrm{low}r}\right) = \mathcal{T}\left(x_{\mathrm{hi}r}\right)|_{\Omega}$. Under this restriction, the estimates constitute an affine subspace $\hat{X}_r$ of codimension $n_0^{\mathrm{h}}n_0^{\mathrm{v}}$ in the $n^{\mathrm{h}}n^{\mathrm{v}}$-dimensional linear space X: $\hat{X}_r = \{\hat{x}_r \in X : \mathcal{T}\left(\hat{x}_r\right)|_{\Omega} = \beta\mathcal{T}_0\left(x_{\mathrm{low}r}\right)\}$.

For $R = 1$, the observation model (1.14) corresponds to the image up-sampling problem. Whenever $R > 1$, we are instead in the image or video super-resolution setting.

1.4.3 Scaling family of transforms

The observation model (1.14) uses a pair of transforms $\mathcal{T}$ and $\mathcal{T}_0$ for relating the same image at two different resolutions. We can extend this to an arbitrary number of intermediate resolutions by means of the following scaling family of transforms.

Let $\{\mathcal{T}_m\}_{m=0}^{M}$ be a family of orthonormal transforms of increasing sizes $n_m^{\mathrm{h}} \times n_m^{\mathrm{v}}$, $n_m^{\mathrm{h}} < n_{m+1}^{\mathrm{h}}$, $n_m^{\mathrm{v}} < n_{m+1}^{\mathrm{v}}$, such that for any pair m,m' with $m < m'$, up to a scaling factor $\beta_{m,m'}$, the whole $\mathcal{T}_m$-spectrum can be considered as a smaller portion of the $\mathcal{T}_{m'}$-spectrum. In particular, this means that the supports Ω_m of the $\mathcal{T}_m$-transform coefficients form a *nested sequence* of subsets (subbands) of Ω_M, i.e. $\Omega_0 \subset \cdots \subset \Omega_M$. The most notable examples of such $\{\mathcal{T}_m\}_{m=0}^{M}$ families are discrete cosine (DCT) or Fourier (DFT) transforms of different sizes, discrete wavelet transforms (DWT) associated to one common scaling function, as well as block DCT, DFT and DWT transforms. Figure 1.9 illustrates the nested sequence of supports for these families.

(a)DFT (b)Block-DCT

(c)DWT,DCT
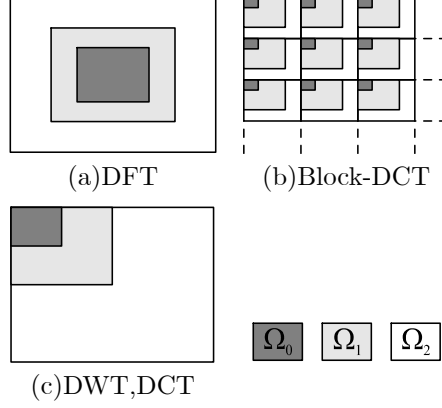
FIGURE 1.9
Nested support subsets: the subsets Ω_0 and Ω_1 are shown as the shaded sub-areas of the support $\Omega_M = \Omega_2$ of coefficients of the transform $\mathcal{T}_M$.

Depending on the specific transform family of choice, the sets Ω_m are commonly referred to as *lower-resolution, low-frequency,* or *coarser-scale subbands* of the $\mathcal{T}_M$-spectrum.

To use such scaling family for the observation model (1.14), we set $n_M^{\mathrm{h}} = n^{\mathrm{h}}$, $n_M^{\mathrm{v}} = n^{\mathrm{v}}$, $\mathcal{T}_M = \mathcal{T}$, $\Omega_0 = \Omega$, $\beta_{0,M} = \beta$, and for $m < m'$ we define the following three operators:

- the *restriction operator* $|_{\Omega_{m,m'}}$ that, from a given $\mathcal{T}_{m'}$-spectrum, extracts its smaller portion defined on Ω_m, which can be thus considered as the $\mathcal{T}_m$-spectrum of a smaller image;

- the *zero-padding operator* $\mathcal{U}_{m,m'}$ that expands a $\mathcal{T}_m$-spectrum defined on Ω_m to the $\mathcal{T}_{m'}$-spectrum defined on the superset $\Omega_{m'} \supset \Omega_m$ by introducing zeros in the complementary $\Omega_{m'} \setminus \Omega_m$;

- the *projection operator* $\mathcal{P}_{m,m'}$ that zeroes out all coefficients of $\mathcal{T}_{m'}$-spectrum defined on Ω_m. If $m = 0$ and $m' = M$, the operator $\mathcal{P}_{m,m'} = \mathcal{P}_{0,M}$ coincides with the projection operator $\mathcal{P}_{\mathrm{ker}(\mathcal{H})}$ defined in Section 1.2.

Note that $\mathcal{U}_{m,m'}(A)|_{\Omega_m} = A$ for any $\mathcal{T}_m$-spectrum A, and $B = \mathcal{P}_{m,m'}(B) + \mathcal{U}_{m,m'}\left(B|_{\Omega_m}\right)$ for any $\mathcal{T}_{m'}$-spectrum B. Thus, $\mathcal{U}_{m,m'}$ can be regarded as "dual" operator of $|_{\Omega_{m,m'}}$.

With this notation, the super-resolution observation model (1.14) becomes

$$y_r = x_{\mathrm{low}r} = \mathcal{T}_0^{-1}\left(\beta_{0,M}^{-1}\,\mathcal{T}_M\left(x_{\mathrm{hi}r}\right)|_{\Omega_{0,M}}\right). \tag{1.15}$$

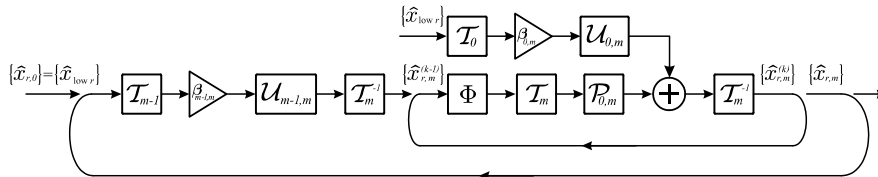### 1.4.4   Multistage iterative reconstruction

In the iterative algorithm presented for the general CS reconstruction, at each iteration the noise excitation and the adaptive filtering are used to provide estimates for the whole spectrum. The algorithm presented below is different, in that it exploits the multiscale feature of the images and performs the scaling from $n_0^{\mathrm{h}} \times n_0^{\mathrm{v}}$ size to $n_M^{\mathrm{h}} \times n_M^{\mathrm{v}}$ gradually, using the transform family $\{\mathcal{T}_m\}_{m=0}^M$ across M *stages*, which are indicated using the subscript $m = 1, \ldots, M$.

The complete algorithm is defined by the iterative system

$$
\begin{cases}
\hat{x}_{r,0} = x_{\mathrm{low}\,r} \\
\hat{x}_{r,m} = \hat{x}_{r,m}^{(k_{\mathrm{final}}(m))} \\
\hat{x}_{r,m}^{(0)} = \mathcal{T}_m^{-1}\left(\mathcal{U}_{m-1,m}\left(\beta_{m-1,m}\mathcal{T}_{m-1}\left(\hat{x}_{r,m-1}\right)\right)\right) \\
\hat{x}_{r,m}^{(k)} = \mathcal{T}_m^{-1}\left(\mathcal{U}_{0,m}\left(\beta_{0,m}\mathcal{T}_0\left(x_{\mathrm{low}\,r}\right)\right) + \mathcal{P}_{0,m}\left(\mathcal{T}_m\left(\Phi\left(r,\left\{\hat{x}_{r,m}^{(k-1)}\right\}_{r=1}^R,\sigma_{k,m}\right)\right)\right)\right).
\end{cases}
$$

$$(1.16)$$

At each stage, the images are being super-resolved from size $n_{m-1}^{\mathrm{h}} \times n_{m-1}^{\mathrm{v}}$ to $n_m^{\mathrm{h}} \times n_m^{\mathrm{v}}$. The sequence $\{x_{\mathrm{low}\,r}\}_{r=1}^R$ serves as input for the first stage, and the output of the current stage $\{\hat{x}_{r,m}\}_{r=1}^R$ becomes an input for the next one. At each stage, the initial estimate $\hat{x}_{r,m}^{(0)}$ is obtained from $\hat{x}_{r,m-1}$ by zero-padding its spectra following the third equation in (1.16). During the subsequent iterations, the estimates are obtained according to the last equation in (1.16), where the superscript $k = 0, 1, 2, \ldots$ corresponds to the iteration count inside each stage, $\hat{x}_{r,m}^{(k)}$ is a sequence of estimates for $\hat{x}_{r,m}$, Φ is a spatially adaptive filter and $\sigma_{k,m}$ is a parameter controlling the strength of this filter. In other words, at each iteration we jointly filter the images $\left\{\hat{x}_{r,m}^{(k-1)}\right\}_{r=1}^R$ obtained from the previous iteration, perform a transform $\mathcal{T}_m$ for each r, substitute the $n_0^{\mathrm{h}} \times n_0^{\mathrm{v}}$ coefficients defined on Ω_0 with $\beta_{0,M}\mathcal{T}_0\left(x_{\mathrm{low}\,r}\right)$, and take an inverse transform $\mathcal{T}_m^{-1}$ to obtain $\hat{x}_{r,m}^{(k)}$. The flowchart of the system (1.16) is presented in Figure 1.10. The iteration process stops at iteration $k_{\mathrm{final}}(m)$ when the distance between $\left\{\hat{x}_{r,m}^{(k)}\right\}$ and $\left\{\hat{x}_{r,m}^{(k-1)}\right\}$ in some metric becomes less than a certain threshold δ_0, or if the maximum number of iterations $k_{\max}(m)$ is reached. In order to prevent smearing of the small details, the sequence $\{\sigma_{k,m}\}_{k=0,1,\ldots}$ should be decreasing with the progress of the iterations.

There are a number of reasons in favor of the multistage reconstruction. Firstly, since at every stage the complexity of each iteration depends on the size $n_m^{\mathrm{h}} \times n_m^{\mathrm{v}}$ of the image to reconstruct, we have that through multistage reconstruction we are able to decrease the computational cost of recovering the coarsest portions of the missing spectrum. Further, the recovery of the finest details becomes more stable, because when we arrive towards the final stages, most of the spectrum has already been reconstructed, with good approximation, in the earlier stages. It turns out that, in this way, the benefit of the stochastic excitation (as in Section CS) is also reduced, in as much as we

FIGURE 1.10
Multistage iterative reconstruction. The inner loop corresponds to an iteration process inside a stage; the outer loop corresponds to a transition to the next stage.

do not include this element in our implementation of the SR reconstruction algorithm (1.12).

1.4.5 Experiments

We present results from three sets of experiments. First, using a synthetic image sequence, we assess how well our SR algorithm can deal with highly aliased data, provided that the set of LR images covers the whole HR sampling grid. Second, we consider video SR reconstruction. Finally, we demonstrate examples of image upsampling showing that successful reconstruction can be obtained even in the case when the exact blurring model is unknown.

As seen in the previous section, the algorithm formulations for upsampling and super-resolution coincide. In both cases, the algorithm performs a reconstruction for each image of the input sequence, and the output sequence always contains the same number of frames as the input one. Whether the algorithm performs upsampling or SR reconstruction, depends on the number of frames in which V-BM3D is allowed to search for similar blocks (so-called "temporal search window"). When the search is restricted to the current frame only, the algorithm independently *upsamples* each frame of the input sequence.

1.4.5.1 Implementation details

The filter's internal 3-D transform is implemented as a composition of a 2-D discrete sine transform (DST) applied to each grouped block and of a 1-D Haar wavelet transform applied along the third dimension of the group. The block size is decreasing with the stages, while within each stage $\sigma_{k,m}$ is decreasing linearly with respect to k. In terms both of smoothing and scale, this consistent with the "coarse-to-fine" approach analyzed in [3]. The temporal search window is set to be equal to the number of frames in the input sequence.

The block-matching in BM3D is implemented as a smallest l_2-difference search. It has been found that strong aliasing in the LR images can noticeably impair the block matching. To overcome this problem, during the first few iterations of the reconstruction algorithm, the block matching is com-

puted over smoothed versions of the current image estimates. Furthermore, the BM3D-based α-rooting described in [11] can be embedded inside V-BM3D so to enhance the high-frequency components during the last iterations. These simple adjustments effectively improve the overall numerical and visual quality of reconstruction, leading to sensibly better results than those reported in [13],[12].

In order to avoid the influence of border distortions and provide more fair comparisons, in all experiments the PSNR values are calculated over the central part of the images, omitting a border of 15-pixel width.
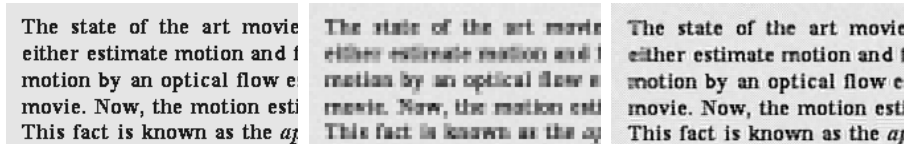
1.4.5.2   Super-resolution

In our super-resolution experiments, we used the same four sequences as in [29] namely: *Text*, *Foreman*, *Suzie* and *Miss America*. Both LR and ground truth HR sequences are available at the website of the first author of [29] http://www.cs.technion.ac.il/~matanpr/NLM-SR/. All LR sequences were obtained using the observation model where the HR images are first blurred using a 3×3 uniform kernel, then decimated by factor 3 and contaminated with additive Gaussian noise with standard-deviation equal to 2. Since our observation model does not assume the presence of noise, we prefilter the noisy LR input sequence with the standard V-BM3D filter using default parameters [9].

A scaling family of transforms can be easily associated to the described observation model, noticing that the LR images can be treated (up to a scaling factor $\beta_{0,M} = 3$) to be composed of DC coefficients of some orthogonal 3×3 block transform. The transform family $\{\mathcal{T}_m\}_{m=0}^{M}$ has been chosen to consist of three 2-D block transforms with 1×1, 2×2, and 3×3 block sizes, which results in a progressive enlargement of 2 and 1.5 times, providing an overall enlargement of 3 times. As a particular family of transforms satisfying the above conditions, we choose the block DCT transforms.

Image super-resolution. For the image super-resolution experiment the *Text* sequence is used. The 9-image LR sequence has been obtained from a single ground-truth image (shown in Figure 1.11(left)) by displacing it before applying observation model described above. The displacements are chosen so that the entire HR grid is covered by the union of the LR grids (i.e. $d^h, d^v = 0, 1, 2,$). One of these nine LR images is shown in Figure 1.11(center), while a super-resolved image obtained by the proposed algorithm is presented in the Figure 1.11. We can see that despite the strong aliasing in the LR images, the algorithm succeeds in reconstructing a readable text.

Video super-resolution. We performed SR of the video sequences *Foreman*, *Suzie* and *Miss America* mentioned above. For a comparison, we also present results obtained by the method [29].

The mean (over all 30 frames) PSNR values for the reconstructed sequences are summarized in Table 1.1. The numerical results obtained by our algorithm are superior to those of [29]. A visual comparison is provided in Figures 1.12 - 1.14. We can observe that while both methods provide roughly the same

FIGURE 1.11
Super-resolution result for the *Text* image. From left to right: original high-resolution image (ground truth); pixel-replicated low-resolution image; image super-resolved by the proposed algorithm.

amount of reconstructed image details, in terms of artifacts, the results of the proposed method are much cleaner.

1.4.5.3   Image upsampling

Let us present also images upsampled of factors $q = 4$ or $q = 8$ from their original resolution. It should be emphasized that in this case we do not know which blurring and decimation operators have been used to obtain the given images. Instead, we assume that the blurring kernel is the low-pass analysis filter of a wavelet transform. Hence, we seek a high-resolution image whose wavelet approximation coefficients in the LL subband of the $\log_2(q)$-level decomposition coincide (up to a scaling factor $\beta_{0,M} = q$) to the pixel values of the given low-resolution image.

Figure 1.15 shows three fragments of the *Cameraman*, *Text*, and *Lighthouse* images at their original resolution. We upsample these fragments applying the $\log_2(q)$-stage algorithm with the Symlet-8 wavelet transform. The obtained high-resolution images are shown in Figure 1.16. It is interesting to notice that the results are quite reasonable, despite our fictitious assumptions about the blurring and decimation operators. The visual quality is particularly good, due to the sharp edges and because of the virtual absence of ringing artefacts typical of transform-domain upsampling.

	Nearest neighbor	[29]	Proposed
Foreman	29.0	32.9	**35.0**
Suzie	30.3	33.0	**34.2**
Miss America	32.0	34.7	**37.0**

TABLE 1.1
Mean (over all frames) PSNR (dB) values of the super-resolved video sequences (see Sec. 1.4.5.2).

1.5   Conclusions

In this chapter we discussed the application of spatially adaptive filters as regularization constraint in inverse imaging problems. Using BM3D as the leading example of such filters, we demonstrated that even simple iterative schemes, when coupled with a good filter, can be turned into powerful and competitive reconstruction methods.

Overall, in the context of compressing sensing, our method introduces a new and alternative view on the reconstruction strategy from undersampled data. In super-resolution applications, the algorithm proposed in Section 1.4 stands in line with the best super-resolution algorithms, possessing registration-free reconstruction, and showing a state-of-the-art performance.

From a general perspective, the presented material expands the breadth of filtering in the modern image processing.
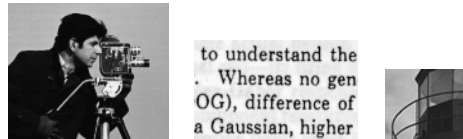
FIGURE 1.12
Results for the 23rd frame from the *Foreman* sequence. From left to right
and from top to bottom: pixel-replicated low-resolution image; original image
(ground truth); super-resolved by the algorithm [29]; super-resolved by the
proposed algorithm and their respective zoomed fragments.

FIGURE 1.13
Results for the 23rd frame from the *Suzie* sequence. From left to right and from top to bottom: pixel-replicated low-resolution image; original image (ground truth); super-resolved by the algorithm [29]; super-resolved by the proposed algorithm and their respective zoomed fragments.

FIGURE 1.14
Results for the 23rd frame from the *Miss America* sequence. From left to right and from top to bottom: pixel-replicated low-resolution image; original image (ground truth); super-resolved by the algorithm [29]; super-resolved by the proposed algorithm and their respective zoomed fragments.

FIGURE 1.15

Fragments of the *Cameraman*, *Text*, and *Lighthouse* images.
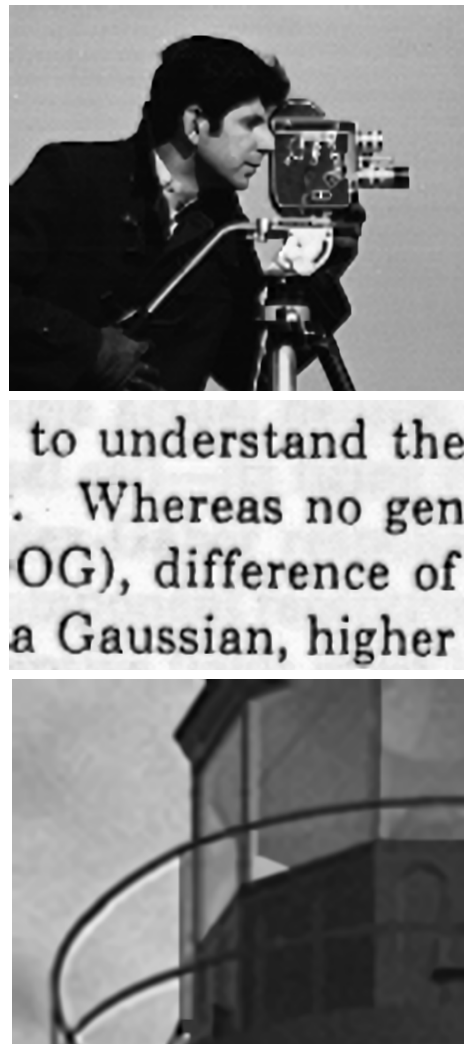
FIGURE 1.16

Upsampling of the three fragments shown in Figure 1.15. From top to bottom: *Cameraman* (4 times), *Text* (4 times), *Lighthouse* (8 times).

Bibliography

[1] R.G. Baraniuk. Compressive sensing [lecture notes]. *Signal Processing Magazine, IEEE*, 24(4):118–121, July 2007.

[2] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation*, 4(2):490–530, 2005.

[3] A. Buades, B. Coll, J.-M. Morel, and C. Sbert. Self-similarity driven color demosaicking. *Image Processing, IEEE Transactions on*, 18(6):1192–1202, June 2009.

[4] A. Buades, B. Coll, and J.M. Morel. Nonlocal image and movie denoising. *Int. J. Computer Vision*, 76(2):123–139, February 2008.

[5] E. J. Candes and J. Romberg. Practical signal recovery from random projections. In *Wavelet Applications in Signal and Image Processing XI, Proc. SPIE Conf.*, volume 5914, 2005.

[6] E. J. Candes and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23:969–985, 2007.

[7] E.J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, Feb. 2006.

[8] J.B. Conway. *A course in functional analysis.* Springer, 1990.

[9] K. Dabov, A. Foi, and K. Egiazarian. Video denoising by sparse 3D transform-domain collaborative filtering. In *Proc. European Signal Process. Conf.*, Poznan, Poland, Sep. 2007.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, Aug. 2007.

[11] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Joint image sharpening and denoising by 3d transform-domain collaborative filtering. In *Proc. 2007 Int. TICSP Workshop Spectral Meth. Multirate Signal Process.*, Moscow, Russia, Sep. 2007.

[12] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian. Image and video super-resolution via spatially adaptive block-matching filtering. In *Proc. Int. Workshop Local Non-Local Approx. Image Process.*, Lausanne, Switzerland, Aug. 2008.

[13] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian. Image upsampling via spatially adaptive block-matching filtering. In *Proc. European Signal Process. Conf.*, Lausanne, Switzerland, Aug. 2008.

[14] D. L. Donoho and M. Elad. Maximal sparsity representation via l_1 minimization. *Proc. Nat. Aca. Sci.*, 100:2197–2202, 2003.

[15] D.L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, Apr. 2006.

[16] M. Ebrahimi and E. R. Vrscay. Multi-frame super-resolution with no explicit motion estimation. In *Proc. Computer Vision and Pattern Recognition*, pages 455–459, Las Vegas, Nevada, USA, July 2008.

[17] K. Egiazarian, A. Foi, and V. Katkovnik. Compressed sensing image reconstruction via recursive spatially adaptive filtering. In *Proc. IEEE Int. Conf. on Image Process., ICIP 2007*, pages 549–552, Sept. 2007.

[18] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14:47–57, 2004.

[19] R. Gribonval and M. Nielsen. Sparse representation in unions of bases. *Information Theory, IEEE Transactions on*, 49(12):3320–3325, 2003.

[20] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola. From local kernel to nonlocal multiple-model image denoising. *International Journal of Computer Vision*, 86(1):1–32, Jan. 2010.

[21] V. Katkovnik and O. Yu. Kulchitskii. Parametric statistical estimates for multiextremum optimization problems. *Engineering Cybernetics*, (6), 1975.

[22] C. Kervrann and J. Boulanger. Local adaptivity to variable smoothness for exemplar-based image denoising and representation. Technical Report 5624, INRIA, July 2005.

[23] C. Kervrann and J. Boulanger. Unsupervised patch-based image regularization and representation. In *Proc. Eur. Conf. Computer Vision (ECCV'06)*, volume 4, pages 555–567, 2006.

[24] C. Kervrann and J. Boulanger. Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *Int. J. Comput. Vision*, 79(1):45–69, 2008.

[25] H. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications, 2nd ed.* Springer, 2003.

[26] S. Lansel. DenoiseLab. http://www.stanford.edu/~slansel/DenoiseLab.

[27] T. Olson. Stabilized inversion for limited angle tomography. *Engineering in Medicine and Biology Magazine, IEEE*, 14(5):612–620, Sep/Oct 1995.

[28] M. Protter and M. Elad. Super resolution with probabilistic motion estimation. *Image Processing, IEEE Transactions on*, 18(8):1899–1904, Aug. 2009.

[29] M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the nonlocal-means to super-resolution reconstruction. *Image Processing, IEEE Transactions on*, 18(1):36–51, Jan. 2009.

[30] H. Takeda, P. Milanfar, M. Protter, and M. Elad. Super-resolution without explicit subpixel motion estimation. *Image Processing, IEEE Transactions on*, 18(9):1958–1975, Sept. 2009.

[31] J.A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *Information Theory, IEEE Transactions on*, 52(3):1030–1051, March 2006.

[32] J.A. Tropp and A.C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666, Dec. 2007.

[33] Y. Tsaig and D. L. Donoho. Extensions of compressed sensing. *Signal Process.*, 86(3):549–571, March 2006.

[34] M.J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ell_1 -constrained quadratic programming (lasso). *Information Theory, IEEE Transactions on*, 55(5):2183–2202, May 2009.