

On the Role of Sparse and Redundant Representations in Image Processing

In image processing, filling in missing portions of images or clearing up blurred images can be rapid, efficient, accurate, and relatively simple procedures.

By MICHAEL ELAD, Senior Member IEEE, MÁRIO A. T. FIGUEIREDO, Fellow IEEE,
AND YI MA, Senior Member IEEE

ABSTRACT | Much of the progress made in image processing in the past decades can be attributed to better modeling of image content and a wise deployment of these models in relevant applications. This path of models spans from the simple ℓ_2 -norm smoothness through robust, thus edge preserving, measures of smoothness (e.g. total variation), and until the very recent models that employ sparse and redundant representations. In this paper, we review the role of this recent model in image processing, its rationale, and models related to it. As it turns out, the field of image processing is one of the main beneficiaries from the recent progress made in the theory and practice of sparse and redundant representations. We discuss ways to employ these tools for various image-processing tasks and present several applications in which state-of-the-art results are obtained.

KEYWORDS | Deconvolution; denoising; dictionary learning; frames; inpainting; redundant dictionaries; sparse representations; superresolution; wavelets

Manuscript received March 24, 2009; revised November 3, 2009; accepted November 4, 2009. Date of publication February 25, 2010; date of current version May 19, 2010. This work was supported in part by the Israel Science Foundation under Grant 1031/08 and by the European Community's FP7-FET program, SMALL project, under Grant 225913. The work of M. Elad was supported by the Israel Science Foundation under Grant 599/08. The work of M. A. T. Figueiredo was supported by the Fundação para a Ciência e Tecnologia, Portuguese Ministry of Science, Technology, and Higher Education. The work of Y. Ma was supported in part by the National Science Foundation and the U.S. Office of Naval Research.

M. Elad is with the Department of Computer Science, The Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: elad@cs.technion.ac.il).

M. A. T. Figueiredo is with the Instituto de Telecomunicações and the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Lisboa 1049-001, Portugal (e-mail: mario.figueiredo@lx.it.pt).

Y. Ma is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: yima@uiuc.edu).

Digital Object Identifier: 10.1109/JPROC.2009.2037655

I. INTRODUCTION

A close inspection of the progress made in the field of image processing in the past several decades reveals that much of it is a direct consequence of the better *image modeling* employed. Armed with a stronger and more reliable model, one can better handle applications ranging from sampling, denoising, restoration, and reconstruction in inverse problems all the way to compression, detection, separation, and beyond. Indeed, the evolution of models for visual data is at the heart of the image-processing literature.

What is a model and why do we need one? We provide an initial answer to these questions through a simple example of noise removal from an image. Given a noisy image, a denoising algorithm is essentially required to separate the noise from the (unknown) clean image. Such a separation clearly requires a close familiarity with the characteristics of both the noise and the original image. Knowing that the noise is additive, white, and Gaussian (AWG) is a good start but far from being sufficient, since the underlying image may also behave like such noise, thereby making the separation of the two impossible. The additional information on the clean image content, that will allow separating it from the AWG noise, constitutes what we refer to in this paper as an *image model*. A classic example of such a model is the intuitive assumption that nearby pixels in “well-behaved” images exhibit strong correlation; i.e., natural images tend to be spatially (piecewise) smooth.

There is a long path of models and their usage in image processing. This path spans from the simple ℓ_2 -norm on local differences (expressing smoothness), through robust and thus edge-preserving measures of smoothness, such as

the total variation [12], [49]. On a different track, it was observed that sparsity of the wavelet coefficients could be used as a reasonable image model [41], [42], [44], [50], [51], and this concept has been used frequently in the past decade. More recently, improved versions of this idea bring us to sparse and redundant representation modeling. Recent work on various image-processing applications indicates that models based on sparse and redundant representations lead to state-of-the-art results and encompass a persuasive potential to this field.

In this paper, we explore the role of models based on sparse and redundant representations and their rationale in image processing. We also relate them to other models that are used successfully in image processing, such as the local principal component analysis (PCA), and example-based techniques. We discuss ways to employ sparse and redundant representations in image-processing tasks in general, and then concentrate on several key applications where this model is shown to lead to state-of-the-art results. More specifically, we discuss image denoising, image inpainting, image deblurring, and superresolution reconstruction. We conclude this paper with a wish-list for this field, describing the research frontiers of this important and challenging arena of work.

II. HISTORICAL PERSPECTIVE ON IMAGE MODELING

We shall start our discussion by motivating the quest for a model for images, and then gradually build a case for a model that is based on sparse and redundant representations.

Consider a family of signals—a set of vectors $\mathcal{X} = \{\mathbf{x}_j, j = 0, 1, 2, \dots\}$, such that $\mathbf{x}_j \in \mathbb{R}^n$. To make our discussion more concrete, assume hereafter that each such signal is a $\sqrt{n} \times \sqrt{n}$ pixels image, representing natural and typical image content, with the corresponding vectors being obtained by lexicographically stacking the image pixels. While such images are very diverse vectors in \mathbb{R}^n , we have no reason to assume that they occupy the entire space. Said more accurately, assuming that the pixels in $\mathbf{x} \in \mathcal{X}$ have values in the range $[0, 1]$, these images do not populate or sample the hypercube $[0, 1]^n \subset \mathbb{R}^n$ uniformly. For example, as mentioned above, spatially smooth images occur much more often than highly nonsmooth and disorganized images. This line of reasoning naturally leads to the Bayesian framework of imposing a probability density function (pdf) on the images—a ‘prior’ distribution $P(\mathbf{x})$ [45]. Priors are extensively used in image processing, serving in inverse problems, compression, anomaly detection, and more, because they provide a systematic way of measuring the probability of an image.

Returning to the image denoising example, consider a given image \mathbf{y} , known to be a noisy version of a clean image \mathbf{x} , contaminated by an additive perturbation \mathbf{v} , i.e., $\mathbf{y} = \mathbf{x} + \mathbf{v}$. Assuming that \mathbf{v} has finite energy, $\|\mathbf{v}\|_2 \leq \epsilon$,

the unknown image \mathbf{x} must be in the sphere $\|\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon$. The optimization problem

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} P(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \quad (1)$$

leads to the most probable image $\hat{\mathbf{x}}$ in this sphere, which is an effective estimate of \mathbf{x} . This way the prior is exploited for solving the denoising problem.

Much effort has been allocated in the image processing community to forming adequate priors. One very common way to construct $P(\mathbf{x})$ is to guess its structure based on intuitive expectations on the data content. For example, the Gibbs distribution $P(\mathbf{x}) \propto \exp\{-\lambda\|\mathbf{L}\mathbf{x}\|_2^2\}$ uses a Laplacian matrix (defined as the linear space-invariant operation that applies the Laplacian filter to the image \mathbf{x}) to give an evaluation of the probability of the image \mathbf{x} . In such a prior, deviation from spatial smoothness, as measured by the Laplacian, is used for judging the probability of the signal. This prior has been extensively used in image processing and is known to be related to both *Tikhonov regularization* and *Wiener filtering* [2]. It is also worth mentioning that this is a particular instance of the more general class of image models known as Markov random fields [32].

The above smoothness prior is known to cause image oversmoothing when used in various image enhancement and restoration tasks. The culprit is the squared ℓ_2 -norm in the exponent of that prior, which strongly penalizes (i.e., makes highly unlikely) any large local differences such as edges, which are key features for visual perception [6]. One remedy for this problem was found to be the replacement of the squared ℓ_2 -norm by a more robust measure, such as the ℓ_1 -norm; by penalizing less any large values, the resulting probability density on $\mathbf{L}\mathbf{x}$ is allowed to have heavy tails. Thus, a prior of the form $P(\mathbf{x}) \propto \exp\{-\lambda\|\mathbf{L}\mathbf{x}\|_1\}$ was found to be more adequate, thus becoming popular in recent years. Similar in spirit is the total-variation prior that also promotes smoothness, but differently, by replacing the Laplacian with gradient norms, thereby using first derivatives rather than second ones. Interestingly, the adoption of the ℓ_1 -norm measure is known to lead to an enforcement of sparsity of the signal/image derivatives.

Another property that can be used for constructing an image prior is sparsity of the wavelet transform, as was experimentally observed in natural images [41], [44], [50], [51]. The orthogonal wavelet transform of an image \mathbf{x} is given by $\mathbf{T}\mathbf{x}$, where \mathbf{T} is a specially designed orthogonal matrix ($\mathbf{T}^T\mathbf{T} = \mathbf{T}\mathbf{T}^T = \mathbf{I}$) containing in its rows spatial derivatives of varying scale, thereby providing what is known as ‘‘multiscale’’ analysis of the signal [42]. Therefore, the prior in this case becomes $P(\mathbf{x}) \propto \exp\{-\lambda\|\mathbf{T}\mathbf{x}\|_1\}$, and this is known to adequately model natural images. Here, the resemblance to the total variation and the Laplacian priors is evident, as in all these cases derivatives of some sort and a robust measure are combined in forming the prior $P(\mathbf{x})$.

One may adopt a common point of view towards these priors, regarding them as mere attempts to describe a random generator machine \mathcal{M} that supposedly generates the signals of interest. However, in all these cases, while the relative¹ evaluation of the probability for a given signal is easy via the $P(\mathbf{x})$ formulas, drawing a random sample from such distributions is often quite difficult. This brings us to sparse and redundant representation modeling, which is a generative model for synthesizing signals according to a prior defined on the coefficients of the representation.

III. SPARSE AND REDUNDANT REPRESENTATION MODELING

A. A Generative Model

Consider the linear system $\mathbf{D}\boldsymbol{\alpha} = \mathbf{x}$, where $\mathbf{D} \in \mathbb{R}^{n \times m}$ and $\boldsymbol{\alpha} \in \mathbb{R}^m$, and interpret it as a way of constructing an image \mathbf{x} . Each of the m columns of \mathbf{D} is a possible image in \mathbb{R}^n , and we refer to these columns as atomic images and to the matrix \mathbf{D} as a dictionary of atoms. One can interpret \mathbf{D} as the *periodic table* of the fundamental elements in the chemistry that describes our images.

The multiplication of \mathbf{D} by a sparse vector $\boldsymbol{\alpha}$ with $k_0 \ll m$ nonzeros produces a linear combination of k_0 atoms with varying weights, generating an image \mathbf{x} . The sparsity of $\boldsymbol{\alpha}$ can be measured by the ℓ_0^0 “norm” $\|\boldsymbol{\alpha}\|_0^0$, which is the limit of ℓ_p^p as $p \rightarrow 0$. In fact, $\lim_{p \rightarrow 0} \|\boldsymbol{\alpha}\|_p^p = |\{i, \alpha_i \neq 0\}|$, the number of nonzero components of $\boldsymbol{\alpha}$. We refer to the vector $\boldsymbol{\alpha}$ that generates \mathbf{x} as its representation, since it describes which atoms and what “portions” thereof are used for its construction. This process of linearly combining atoms to form an image (think of it as a molecule in the signal chemistry) may be referred to as *atom composition*.

Consider the set of all the possible placements of $k_0 \ll m$ nonzeros, which has cardinality $\binom{m}{k_0} \approx m^{k_0}$. Assume that samples from this set are drawn with uniform probability. Assume further that, given their positions, each nonzero entry in $\boldsymbol{\alpha}$ is drawn independently from the zero-mean τ -variance Gaussian distribution $\mathcal{N}(0, \tau)$. This gives us a complete definition of the pdf of $\boldsymbol{\alpha}$, as well as a probabilistic generative model for images \mathbf{x} —this constitutes the random signal generator $\mathcal{M}(\mathbf{D}, k_0, \tau)$. Note that the image \mathbf{x} is the outcome of a mixture of Gaussians, each of dimension k_0 , and with equal probabilities.

An important addition to the definition of the model \mathcal{M} could be postulating a random perturbation (noise) vector $\mathbf{e} \in \mathbb{R}^n$ with bounded energy, $\|\mathbf{e}\|_2 \leq \epsilon$, such that $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{e}$. Such an additive perturbation may be useful because the model $\mathcal{M}(\mathbf{D}, k_0, \tau)$ imposes too strong a restriction on the images, and using it in practical applications would necessarily lead to a mismatch between

actual observed images and their imposed model. With the introduction of this perturbation, which enables such a mismatch, we refer hereafter to the model proposed as $\mathcal{M}(\mathbf{D}, k_0, \tau, \epsilon)$. We should mention that this model is simplistic, and in later sections we shall expand it to better handle true image data.

B. The Basics on Using this Model

How do we practice image processing with sparse and redundant representation modeling? Suppose we have an image \mathbf{x} , assumed to have been generated by the model $\mathcal{M}(\mathbf{D}, k_0, \tau, \epsilon)$, and that the parameters of the model are known. There are numerous image-processing tasks that could be of interest to us. As shown next, common to these is the need to solve a problem we refer to hereafter as $\mathcal{P}_0(\mathbf{D}, \mathbf{x}, \delta)$, which has the form

$$\mathcal{P}_0(\mathbf{D}, \mathbf{x}, \delta) : \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0^0 \quad \text{subject to } \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2 \leq \delta. \quad (2)$$

Solving this problem corresponds to finding the sparsest vector $\boldsymbol{\alpha}$ that explains \mathbf{x} as a linear combination of columns from \mathbf{D} , with weight vector $\boldsymbol{\alpha}$, with an error no larger than δ . This process is known as *atomic decomposition* [13].

How would one go about solving $\mathcal{P}_0(\mathbf{D}, \mathbf{x}, \delta)$? A direct solution is prohibitive, as it requires a combinatorial search of exponential size for considering all possible supports. Various practical alternative techniques for approximating the solution of this problem have been proposed in recent years. Some rely on a relaxation that replaces $\|\boldsymbol{\alpha}\|_0^0$ by $\|\boldsymbol{\alpha}\|_1$, yielding the problem $\mathcal{P}_1(\mathbf{D}, \mathbf{x}, \delta)$

$$\mathcal{P}_1(\mathbf{D}, \mathbf{x}, \delta) : \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad \text{subject to } \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2 \leq \delta. \quad (3)$$

This formulation may also be interpreted as attributing a Laplace-prior on the representation coefficients.

Another class of methods adopts greedy schemes in which the unknown support is obtained sequentially; detailed descriptions and references may be found in [7]. The important thing to mention in this respect is the fact that a thorough theoretical study leads to guarantees for such approximation methods to perform well (see [7] and the many references therein), thus making the solution of $\mathcal{P}_0(\mathbf{D}, \mathbf{x}, \delta)$ accessible.

Returning to image-processing tasks, here are few core examples where sparse and redundant representation modeling is used for handling commonly encountered problems.

- *Compression*: Nominally, \mathbf{x} requires a description by n numbers. However, if we can solve $\mathcal{P}_0(\mathbf{D}, \mathbf{x}, \delta)$, for some $\delta \geq \epsilon$, then the resulting solution,

¹Due to the need for the normalization coefficient.

denoted as α_0^δ , affords an approximation $\hat{\mathbf{x}} = \mathbf{D}\alpha_0^\delta$ to \mathbf{x} using at most k_0 scalars, with an approximation error at most δ . Notice that if \mathbf{x} was generated by $\mathcal{M}(\mathbf{D}, k_0, \tau, \epsilon)$, it is guaranteed that there exists a $\mathbf{x}_0 = \mathbf{D}\alpha_0$, where $\|\alpha_0\|_0^0 = k_0$, and such that $\|\mathbf{x} - \mathbf{x}_0\|_2 \leq \epsilon \leq \delta$. By increasing δ , we obtain a deeper compression with fewer nonzeros and a larger approximation error. This way (and ignoring quantization issues), we obtain a rate-distortion curve for a compression mechanism.

- *Denoising and Linear Inverse Problems:* Consider a noisy indirect measurement of \mathbf{x} , i.e., $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$, where \mathbf{v} is an additive noise known to obey $\|\mathbf{v}\|_2 \leq \delta$. The operator \mathbf{H} could be the identity (in which case the problem reduces to denoising), it could represent a blur, a tomographic projection, the masking of some pixels (which leads to inpainting), down-sampling (which leads to super-resolution), a random set of projections (which leads to compressed sensing; see [11] and references therein), or any other kind of linear degradation. By solving $\mathcal{P}_0(\mathbf{H}\mathbf{D}, \mathbf{y}, \epsilon + \delta)$, the resulting solution $\alpha_0^{\delta+\epsilon}$ will have at most k_0 nonzeros. A theoretical analysis establishes that if k_0 is small enough, then $\alpha_0^{\delta+\epsilon}$ is at most $\mathcal{O}((\epsilon + \delta)k_0/n)$ away from the original representation α , implying a very effective reconstruction [10]. Thus, we can expect to identify directly the sparse components of the underlying signal \mathbf{x} and obtain an approximation/estimate $\hat{\mathbf{x}} = \mathbf{D}\alpha_0^{\delta+\epsilon}$.
- *Morphological Component Analysis (MCA):* Suppose that the observed signal is a superposition of two different signals, i.e., $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where \mathbf{x}_1 is sparsely generated using model $\mathcal{M}(\mathbf{D}_1, k_1, \tau, \epsilon_1)$ and \mathbf{x}_2 is sparsely generated using model $\mathcal{M}(\mathbf{D}_2, k_2, \tau, \epsilon_2)$. Can we separate the two superimposed signals? For example, can we separate the texture and cartoon contents of an image, given that both are sparsely generated by two quite different dictionaries? Observe that \mathbf{x} can be modeled as having been generated by $\mathcal{M}([\mathbf{D}_1, \mathbf{D}_2], k_1 + k_2, \tau, \epsilon_1 + \epsilon_2)$, where $[\mathbf{D}_1, \mathbf{D}_2]$ denotes the concatenation of the two dictionaries. Thus it makes sense to solve the problem $\mathcal{P}_0([\mathbf{D}_1, \mathbf{D}_2], \mathbf{x}, \epsilon_1 + \epsilon_2)$; the solution $\alpha = (\alpha_1, \alpha_2)$ allows generating a plausible answer to the separation problem: $\hat{\mathbf{x}}_1 = \mathbf{D}_1\alpha_1$ and $\hat{\mathbf{x}}_2 = \mathbf{D}_2\alpha_2$ [27].

A wide range of other applications can also be envisioned. All these applications call for the solution of $\mathcal{P}_0(\mathbf{D}, \mathbf{x}, \delta)$ or variants of it; as mentioned above, these can be approximated reliably by practical algorithms.

C. Geometric Interpretation and Relation to Other Models

In order to gain a geometric insight into the proposed model and some alternatives to it, we return to the begin-

ning of this section, considering a large corpus of examples of images (or image patches) $\mathcal{X} = \{\mathbf{x}_j, j = 0, 1, 2, \dots\} \subset \mathbb{R}^n$. We concentrate on one arbitrary image $\mathbf{x}_0 \in \mathcal{X}$ and its δ -neighborhood, and aim to study the behavior of this neighborhood in \mathbb{R}^n .

For a small enough value of δ , moving δ away from \mathbf{x}_0 along directions $\mathbf{e} = \mathbf{x} - \mathbf{x}_0$, where $\mathbf{x} \in \mathcal{X}$, represents small permissible perturbation directions that lead to feasible signals. The question we pose is: do those fill the entire δ -ball in \mathbb{R}^n space? We denote this set of permissible directions as $\Omega_{\mathbf{x}_0}^\delta$. Gathering all vectors $\mathbf{e} \in \Omega_{\mathbf{x}_0}^\delta$ into an $n \times |\Omega_{\mathbf{x}_0}^\delta|$ matrix $\mathbf{E}_{\mathbf{x}_0}$, we aim to study the behavior of its singular values. We will be interested in families of structured images, for which the effective rank of such matrices is $k_{\mathbf{x}_0} \ll n$, for any $\mathbf{x}_0 \in \mathcal{X}$. This is equivalent to the statement that the $k_{\mathbf{x}_0} + 1$ th singular-value and beyond in these matrices tend to be near-zero and are therefore negligible, and the effective directions in the obtained subspace can be accumulated as an orthogonal set of $k_{\mathbf{x}_0}$ columns in a $n \times k_{\mathbf{x}_0}$ matrix $\mathbf{Q}_{\mathbf{x}_0}$.

Signals satisfying the above local low-dimensionality assumption essentially exhibit a local behavior that is approximately a linear subspace of dimension $k_{\mathbf{x}_0}$, shifted around \mathbf{x}_0 . While this subspace and its dimension may vary from one point to another, all of them are characterized as being far smaller than the ambient n dimensions [35]. Experimental studies show that most informative signals we work with, such as images, audio, and more, follow such a structure, and this can be exploited in processing them [38], [39], [56].

Based on the above description, it is tempting to build a model of a signal source by holding many instances of example signals \mathcal{X} and using these directly to characterize these local subspaces; this is known as a local PCA modeling [56]. In this approach, the columns of $\mathbf{Q}_{\mathbf{x}_0}$ are the principal directions of the signal for the location \mathbf{x}_0 , and those are permitted to vary freely as a function of \mathbf{x}_0 . In general, a complete model of this sort requires either a storage of \mathcal{X} and extraction of $\mathbf{Q}_{\mathbf{x}_0}$ per need or a storage of all the possible unitary bases $\mathbf{Q}_{\mathbf{x}}$ gathered offline for every $\mathbf{x} \in \mathcal{X}$. Both options are prohibitive in most cases.

If we can further assume that a small number of such matrices $\{\mathbf{Q}_p, p = 1, \dots, P\}$ (say, a few hundreds or thousands) covers all possible cases, this model can be made more efficient. This calls for a (fuzzy) clustering of the signals $\mathbf{x} \in \mathcal{X}$ to subgroups that correspond each to a different matrix \mathbf{Q}_p . As the range of each matrix \mathbf{Q}_p is a low-dimensional subspace, we essentially use an arrangement of many low-dimensional subspaces to approximate the distribution of all signals \mathcal{X} . Such a model has been proposed in recent years as a generalization to principal component analysis [56] and has shown to be very effective in grasping the behavior of image content, by working on small patches [35], [38], [39].

The sparse and redundant representation model is one step further in an attempt to make such local model

more concise. Assuming for simplicity that all these local subspaces are of the same dimensionality k_0 , using sparse representations of cardinality k_0 over a dictionary \mathbf{D} with m columns, we have at our disposal $\binom{m}{k_0}$ subspaces, just by allowing k_0 -sparse representations with all the possible supports. This way, our dictionary holds very efficiently all the key directions required to represent the signal anywhere in \mathbb{R}^n while enabling a very large number of possibilities. Clearly, though, this comes at the price of further restricting the structure of the model by assuming that the atoms in the dictionary are principal directions shared by many subspaces.

For completeness of this presentation, we mention that rather than describe the model of the signal by constructing parameters that characterize its local behavior, one could use the local neighbors directly. This is known as a direct example-based modeling, and its use for texture synthesis [21], inpainting, denoising [8], and other applications show promising results. Such a direct technique is very appealing, but it works well only if the sampling of the signal source is dense enough. This approach could be considered as an extreme sparse representation modeling of the signal \mathbf{x} , such that the δ -local neighborhood serves as its local (and varying) dictionary. A related approach models image patches as being probably close to a low-dimensional manifold [48].

Before turning to actual image-processing tasks, we should mention that the model we have denoted as $\mathcal{M}(\mathbf{D}, k_0, \tau, \epsilon)$ is overly simplistic, and one has to expand it to better handle true image data. As we shall see next, we base our actual models on the same foundations as described above, with small variations.

IV. IMAGE-PROCESSING APPLICATIONS

We now turn to discuss specific applications where sparse and redundant representation modeling is shown to be highly effective. We should note that in migrating from the above core ideas to applications, there is much room for imagination and creativity, as a direct deployment of the outlined ideas will not necessarily operate well. Thus, as we show next, each of the applications described below takes a different route in using the model proposed, choosing the dictionary to work with, adopting the numerical scheme to approximate sparse representation, and more.

A. Image Denoising

We obtain an observed image \mathbf{y} , a noisy version of an unknown underlying clean image \mathbf{x} , i.e., $\mathbf{y} = \mathbf{x} + \mathbf{v}$, and our goal is to recover \mathbf{x} . A popular noise model is to consider $\mathbf{v} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. In line with the discussion in Section III-B, denoising could be performed by solving $\mathcal{P}_0(\mathbf{D}, \mathbf{y}, \delta)$ for an appropriate choice of the dictionary \mathbf{D} and the scalar δ . In the following, we describe several such options, as practiced in the past decade. Note that we can

replace $\mathcal{P}_0(\mathbf{D}, \mathbf{y}, \delta)$ and $\mathcal{P}_1(\mathbf{D}, \mathbf{y}, \delta)$ with the equivalent (Lagrangian) form

$$\mathcal{G}_p(\mathbf{D}, \mathbf{y}, \lambda) : \min_{\boldsymbol{\alpha}} \lambda \|\boldsymbol{\alpha}\|_p^p + \frac{1}{2} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \quad (4)$$

that replaces the constraint by a penalty. We shall use these two modes interchangeably, and often times we shall prefer to work with $p > 0$ for obtaining better denoising performance.

1) *Orthogonal Wavelet Denoising*: Wavelet-based models had a strong impact on the field of image processing, especially in coding and denoising. Their success is due to the tendency of images to become sparse in the wavelet transform domain [42], [44], [51]. This implies that image approximations based on a small subset of wavelets are typically very accurate, which is a key to wavelet-based denoising and compression. Interestingly, it has been found that the human visual system exploits this sparseness property by using wavelet-like representations in the visual cortex [36], [47].

As already mentioned in Section II, the discrete wavelet transform (DWT) of an image \mathbf{x} is represented by the multiplication $\boldsymbol{\alpha} = \mathbf{T}\mathbf{x}$, where \mathbf{T} is an orthogonal matrix, i.e., $\mathbf{x} = \mathbf{T}^T\boldsymbol{\alpha}$. Therefore, the dictionary to be used within $\mathcal{G}_p(\mathbf{D}, \mathbf{y}, \lambda)$ in (4) is $\mathbf{D} = \mathbf{T}^T$. Due to the unitarity of \mathbf{T} , we have that $\|\mathbf{D}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 = \|\mathbf{T}^T\boldsymbol{\alpha} - \mathbf{y}\|_2^2 = \|\boldsymbol{\alpha} - \mathbf{T}\mathbf{y}\|_2^2$, and thus our denoising process translates into

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \lambda \|\boldsymbol{\alpha}\|_p^p + \frac{1}{2} \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2 \quad (5)$$

where $\boldsymbol{\beta} = \mathbf{T}\mathbf{y}$ is the DWT of the noisy image; that is, we have problem $\mathcal{G}_p(\mathbf{I}, \boldsymbol{\beta}, \lambda)$. Of course, the final image estimate is obtained as $\hat{\mathbf{x}} = \mathbf{T}^T\hat{\boldsymbol{\alpha}}$. As both terms $\|\boldsymbol{\alpha}\|_p^p$ and $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2^2$ are separable, the optimization decouples into a set of independent scalar problems of the form $\hat{x}_i = \arg \min_x 0.5(x - y)^2 + \lambda|x|^p$, which have particularly simple closed-form solutions in the two notable cases $p = 0$ and $p = 1$, called hard- and soft-thresholding, respectively. The corresponding functions $\text{hard}(y, a)$ and $\text{soft}(y, a)$ are shown in Fig. 1; soft thresholding should use $a = \lambda$, whereas hard thresholding should use $a = \sqrt{2\lambda}$.

The solution of (5) is thus obtained by applying this soft/hard thresholding function in a component-wise fashion to the observed coefficients $\mathbf{T}\mathbf{y}$ and this way obtain $\hat{\boldsymbol{\alpha}}$. Then, an inverse DWT leads to the denoised image $\hat{\mathbf{x}} = \mathbf{T}^T\hat{\boldsymbol{\alpha}}$. This signal/image denoising method just described is the well-known transform, shrink, inverse transform approach, which sparked the explosion of interest in wavelet-based signal/image processing that took place in the early 1990s [18], [19], [43], [44], [51], [55].

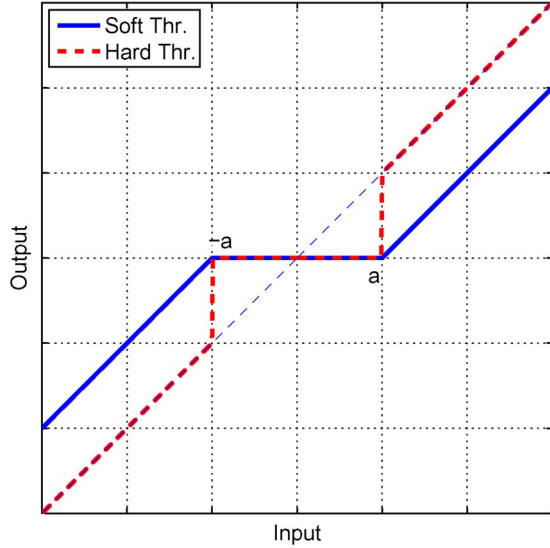


Fig. 1. The soft and hard shrinkage curves.

2) *Denoising With Wavelet Frames*: It is well known that the dyadic image partitioning underlying the orthogonal DWT causes blocky artifacts in the processed images. In denoising problems, translation-invariant approaches have been shown to significantly reduce these artifacts and produce a substantial improvement in the denoising outcome [14], [28], [37].

Let \mathbf{x} be represented with a redundant dictionary \mathbf{D} that has more columns than rows, as advocated in Section III, i.e., $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$. Assume that \mathbf{D} is associated with some transform matrix $\mathbf{T} = \mathbf{D}^+$. This allows computing $\boldsymbol{\alpha}$ from \mathbf{x} as $\boldsymbol{\alpha} = \mathbf{T}\mathbf{x}$, via the well-known perfect reconstruction property $\mathbf{D}\mathbf{T} = \mathbf{I}$ [54]. When \mathbf{D} is a tight frame, we simply have $\mathbf{T} = \mathbf{D}^T$, thus $\mathbf{D}\mathbf{D}^T = \mathbf{I}$; this is the case, for example, for any orthogonal basis. In the case of an overcomplete tight frame, $\mathbf{D}\mathbf{D}^T = \mathbf{I}$ holds but, unlike in an orthogonal basis, $\mathbf{D}^T\mathbf{D} \neq \mathbf{I}$.

A standard example of a tight wavelet frame, known as a shift-invariant representation, contains all possible shifts of each element of an orthogonal wavelet basis. The size of \mathbf{D} in such a case is $n \times (n \log n)$ in the case of a full wavelet decomposition (all the available scales). In recent years, other wavelet-type frames have been introduced, such as *curvelets*, to address the problem of finding optimally sparse representations for images with discontinuities along piecewise smooth edges [9].

Returning to our denoising goal, we adopt again the formulation of $\mathcal{G}_p(\mathbf{D}, \mathbf{y}, \lambda)$ in (4) as a way to derive the denoising algorithm. Thus, we aim to solve

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \lambda \|\boldsymbol{\alpha}\|_p^p + \frac{1}{2} \|\mathbf{D}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \quad (6)$$

with \mathbf{D} as described above.

Unlike (5), this problem does not have simple closed-form solutions, not even in the two notable cases $p = 0$ and $p = 1$ mentioned above, because the presence of \mathbf{D} destroys the separability that was present in (5). The denoising problem posed in (6) can be solved (at least approximately) by greedy techniques if the dimensions involved are relatively low ($n \leq 1000$) [7]. Considering a high-dimensional case (e.g., $n = 10^6$ representing an image of size 1000×1000), such greedy methods are no longer relevant. In recent years, *iterative shrinkage/thresholding* (IST) algorithms, tailored for objective functions of the form (6), were independently proposed by several authors [4], [17], [23], [25], [29], [30], [46], [52], [53]. Recently, the work reported in [15] brought important contributions (including strong convergence results) to the study of a class of *forward-backward splitting* algorithms, which includes IST as a particular member. The IST algorithm for (6) has the form

$$\hat{\boldsymbol{\alpha}}_{t+1} = \arg \min_{\boldsymbol{\alpha}} \lambda \|\boldsymbol{\alpha}\|_p^p + \frac{1}{2} \|\boldsymbol{\alpha} - \boldsymbol{\beta}_t\|_2^2 \quad (7)$$

where $\boldsymbol{\beta}_t = \hat{\boldsymbol{\alpha}}_t - \nu \mathbf{D}^T(\mathbf{D}\hat{\boldsymbol{\alpha}}_t - \mathbf{y})$. Each step of (7) is a pure denoising problem of the form (5), which has a simple closed-form solution. This way, the overall denoising process requires a sequence of simple steps of multiplication by \mathbf{D} and its adjoint, and a scalar shrinkage step. Assuming that the multiplication by the dictionary (and its adjoint) has a fast $O(n \log n)$ algorithm, the overall process is very fast and effective. We should also mention that fast IST-like algorithms were recently proposed by several authors [3], [5], [22], [57].

3) *Denoising With Learned Dictionaries*: To complete our discussion on image denoising, we now turn to a more recent development that employs a learned dictionary \mathbf{D} . Rather than working on the image as a whole, we now adopt the sparse and redundant representation model on small image patches of size $\sqrt{n} \times \sqrt{n}$ (a typical choice is in the range $5 \leq \sqrt{n} \leq 20$) [24]. We assume that each and every patch in the given image (with overlaps) is expected to have a sparse representation with respect to \mathbf{D} . Thus, the dictionary is a small matrix of size $n \times m$ where, for example, $n \approx 100$ and $m \approx 200$.

Embarking from the formulation of $\mathcal{P}_0(\mathbf{D}, \mathbf{y}, \delta)$ in (2), we define the following optimality criterion for denoising the complete image based on the patch model

$$\begin{aligned} \min_{\mathbf{x}, \{\boldsymbol{\alpha}_i\}_{i \in \Omega}} \quad & \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \sum_{i \in \Omega} \|\boldsymbol{\alpha}_i\|_0^0 \\ \text{subject to} \quad & \|\mathbf{R}_i \mathbf{x} - \mathbf{D}\boldsymbol{\alpha}_i\|_2 \leq \delta, \quad \forall i \in \Omega. \end{aligned} \quad (8)$$

In this formulation, the domain of the image is defined as Ω , and we index locations in it by $i \in \Omega$. The operator \mathbf{R}_i

extracts a patch of size $\sqrt{n} \times \sqrt{n}$ from location i . For each patch, we construct a representation α_i that should be both sparse and represent $R_i x$ to within a prespecified error.

Our goal is to find both the set of representations and a clean image estimate \hat{x} . A block-coordinate relaxation technique can be used, where we fix $x = y$ and find the representations $\{\alpha_i\}_{i \in \Omega}$ first. These are found by solving a set of problems of the form $\mathcal{P}_0(D, R_i y, \delta)$. Since these problems are of low dimension, a greedy approach, such as the OMP, can be used effectively. Once found, \hat{x} can be computed by fixing these representations and solving

$$\min_x \frac{1}{2} \|x - y\|_2^2 + \lambda \sum_{i \in \Omega} \|R_i x - D \alpha_i\|_2^2. \quad (9)$$

Notice that the constraints have been turned into a penalty for this expression to be easy to handle. Since this is a purely quadratic expression, its minimization is easily obtained as a linear set of equations. As it so happens, this set of equations leads to a simple averaging of the cleaned patches, with a small portion the noisy image [24].

As for the identity of D , one could use a prespecified dictionary, such as a redundant DCT, or gather many patch examples and learn a dictionary that sparsifies them, with the hope that this property generalizes well to the image to be cleaned [1]. Alternatively, somewhat better denoising can be obtained if the learning procedure is employed on patches extracted from the noisy image y itself, adapting to the specific image content to be handled [24]. Fig. 2 pre-

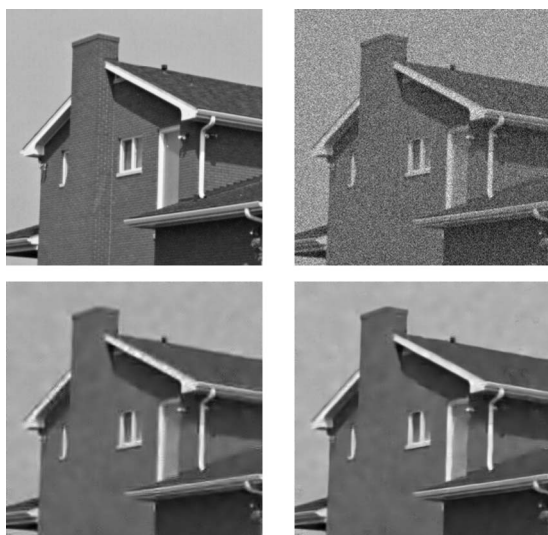


Fig. 2. Denoising results using the local-patch processing algorithm [24]. (Top left) The original image; (top right) the noisy image with $\sigma = 25$ (PSNR = 20.19 dB); (bottom left) denoised using a redundant DCT dictionary (PSNR = 31.00 dB); (bottom right) denoised using an adaptive dictionary (PSNR = 32.25 dB).

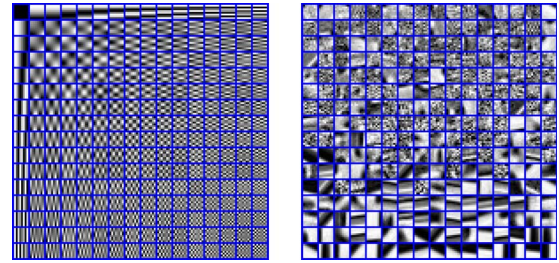


Fig. 3. The atoms of the (redundant) DCT and the adaptive dictionaries. In both cases, the atoms are of size 8×8 pixels, and there are $m = 256$ of them.

sents the denoising results for the DCT and the adaptive dictionaries, and Fig. 3 shows the dictionaries themselves.

B. Image Deblurring

Image deblurring is one of the earliest and most classical linear inverse problems in imaging, dating back to the 1960s [2]. In image deblurring, a noisy blurred version of x is observed, i.e., $y = Hx + v$, where H represents a blur operation, which in practice may result from physical mechanisms such as relative motion between the camera and the subject (motion blur), lack of focus (defocusing blur), or a number of other mechanisms that are well modeled by a convolution.

There are (too) many image-deblurring algorithms, and some of these exploit sparse representation modeling in various ways. We present one such approach here, to illustrate the power of this model, and the simplicity with which one can get state-of-the-art results for this application by a simple adoption of the paradigm presented.

As in the denoising case discussed above, image deblurring can be approached using the form $\mathcal{G}_p(HD, y, \lambda)$ as defined in (4), where we have adopted the idea presented in Section III-B of adding the blur operation into the effective dictionary. Thus the problem that needs to be solved is

$$\hat{\alpha} = \arg \min_{\alpha} \lambda \|\alpha\|_p^p + \frac{1}{2} \|HD\alpha - y\|_2^2. \quad (10)$$

Assuming that we operate on the whole image using a tight frame of some sort (e.g., redundant Haar as in [28]), as described in Section IV-A2, we can apply the IST algorithm to this problem, leading to update equations of the same form as (7) but with HD replacing D . Fig. 4 exemplifies wavelet-based deblurring, where H represents a 9×9 uniform blur and D is a shift-invariant Haar frame.

C. Image Inpainting

Image inpainting refers to the desire to fill in missing values in an image, based on their surrounding neighbors.

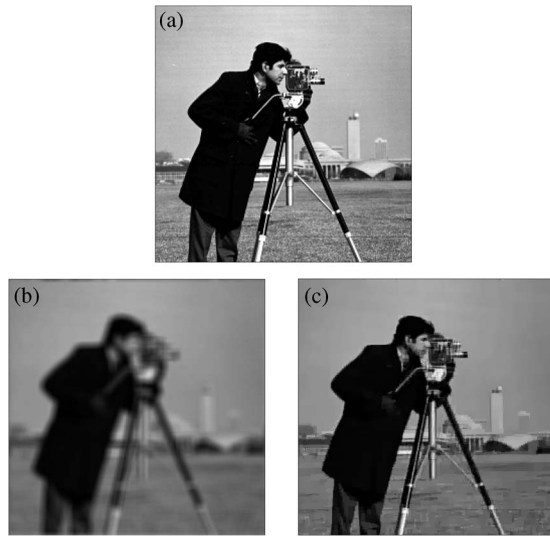


Fig. 4. (a) Original image, (b) blurred image (uniform 9×9), and (c) deblurred image, using the IST algorithm.

Such a problem is often encountered in a faulty transmission of image content over unreliable channels, when missing blocks are to be recovered. Another instance of the inpainting problem is obtained in cases of scratches in images (e.g., old film) that are to be fixed, or if one desires to manipulate an image content to remove portions from it. Ever since its introduction in the image-processing literature, this problem has attracted considerable attention, and many solutions have been proposed. One popular family of inpainting algorithms is based on partial differential equations that propagate the information from the borders of the holes in the image.

More recently, sparse representation modeling found its way into this field, with promising performance. At its core, the inpainting problem can be formulated as the problem $\mathcal{G}_p(\mathbf{M}\mathbf{D}, \mathbf{y}, \lambda)$, where \mathbf{M} is a diagonal mask matrix of size $n \times n$, with ones for existing pixels and zeroes elsewhere. We further assume that in the missing pixels, the image \mathbf{y} is set to zero. Having posed the problem this way, all the above discussion on denoising and deblurring becomes relevant, and in fact leads to an effective solution of the problem.

There are several algorithms proposed for the inpainting problem along the above lines, some of which adopt the global approach of handling the image as a whole [27] and others that operate on patches [33], [34], [40]. Due to their close resemblance to deblurring algorithms, we shall not dwell further on this topic and simply show typical examples. The results shown in Fig. 5 are taken from [27]. This inpainting algorithm follows the global approach and combines MCA (for separation of cartoon and texture content) with inpainting. The MCA dictionary used is composed of curvelets and local DCT; as can be seen, the quality of the filling-in is near perfect.

D. Superresolution

With high-definition TV and electronic displays' becoming ever more popular, a very pressing task is how to convert all the old images and videos to a higher resolution, such that their sharpness and naturalness match those of true high-resolution images. Image superresolution is arguably one of the most classical inverse problems in image processing and is, by nature, intrinsically underdetermined. The problem can be simply stated as that of recovering a high-resolution image $\mathbf{x} \in \mathbb{R}^n$ from its low-resolution version $\mathbf{y} \in \mathbb{R}^k$ (with $k < n$). We model the relation between these two by

$$\mathbf{y} = \mathbf{S}\mathbf{H}\mathbf{x} = \mathbf{L}\mathbf{x} \quad (11)$$

where \mathbf{H} is a linear filter that models certain low-pass filtering (blurring, e.g., with a Gaussian kernel), \mathbf{S} is a down-sampling operator, and $\mathbf{L} = \mathbf{S}\mathbf{H}$. The dimension of \mathbf{y} is significantly smaller than that of \mathbf{x} ; thus there are infinitely many possible vectors \mathbf{x} that satisfy the above equation.

Obviously, to obtain a unique and "good" high-resolution image, proper regularization is needed by imposing certain priors on the solution. As discussed in Section III-C, one plausible model for natural images is based on local PCA. The model suggests that a small image patch is likely to have a very sparse representation with respect to a dictionary of densely sampled patches from natural images. Empirical evidence has suggested that a large set (of the order of a hundred thousand) of patches randomly sampled from natural images is in fact an excellent candidate for such a dictionary [59].

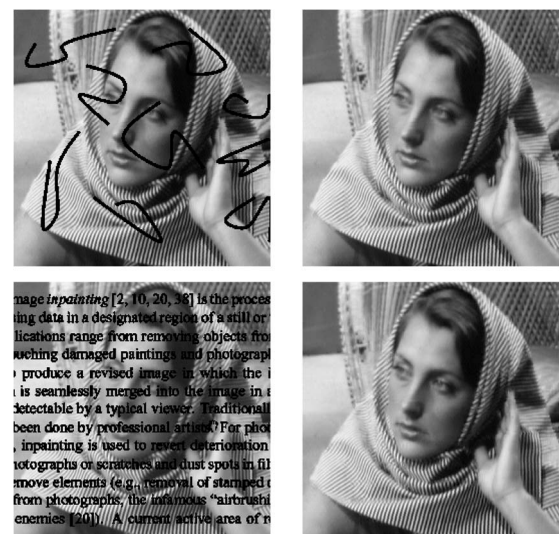


Fig. 5. Two input images with missing pixels (in black) are given on the left (top and bottom). The corresponding MCA inpainting results from [27] are shown on the right.



Fig. 6. The girl image magnified by a factor of four. (Top left) Low-resolution input. (Top right) ℓ^2 -based back-projection. (Bottom left) Learning-based method in [31]. (Bottom right) The sparse representation method.

With respect to such a dictionary of high-resolution patches, denoted as \mathbf{D}_h , we may assume that any high-resolution image patch has a sparse representation

$$\mathbf{x} = \mathbf{D}_h \boldsymbol{\alpha} \quad (12)$$

for some $\boldsymbol{\alpha}$ with $\|\boldsymbol{\alpha}\|_0 \leq k$. Thus, the superresolution problem becomes that of recovering $\boldsymbol{\alpha}$ from low-dimensional measurements

$$\mathbf{y} = \mathbf{Lx} = \mathbf{LD}_h \boldsymbol{\alpha}. \quad (13)$$

Notice that $\mathbf{D}_l = \mathbf{LD}_h$ can be seen as a dictionary of corresponding low-resolution image patches. We may thus attempt to recover $\boldsymbol{\alpha}$ by solving $\mathcal{P}_p(\mathbf{D}_l, \mathbf{y}, \delta)$ or $\mathcal{G}_p(\mathbf{D}_l, \mathbf{y}, \lambda)$. This seemingly naive choice of dictionary has turned out to be extremely effective for generating high-quality high-resolution images. Such a sparsity-based approach for superresolution has been thoroughly explored in [58] and [59]. Fig. 6 shows a typical result in comparison with

other state-of-the-art superresolution methods in the literature. It is striking to see that such a simple scheme achieves qualitatively and quantitatively better results than other much more sophisticated learning or estimation schemes.

V. SUMMARY AND CONCLUSION

In this paper, we have briefly reviewed sparse and redundant representations as a new model that harnesses the local low-dimensional structure of natural images. Overwhelming theoretical and empirical evidence suggests that this model may provide one possibility for a unified mathematical and computational framework for image processing. Within this framework, one can develop more efficient and effective solutions to many conventional image-processing tasks, including but not limited to image compression, denoising, deblurring, inpainting, superresolution, segmentation, and more.

Despite its success so far, many difficult problems remain open at this stage. Topics such as general-purpose image compression with this model, use of multiscale learned dictionaries and their deployment in applications, modifications of the model to better fit images, synthesis of images based on sparse representations, handling geometrical structures, and more are all questions unanswered as of yet. Indeed, put more broadly, the very fundamental question of why this model suits images so well, and why algorithms evolving from it operate successfully in some applications, are also questions to be discussed. These and more form a major part of a wish-list for the future research in this field. We hope this paper inspires researchers to investigate these problems further and gain better insights about image modeling and processing. ■

Acknowledgment

M. Elad thanks M. Aharon, D. L. Donoho, and J.-L. Starck, with whom he worked on image denoising and inpainting. M. A. T. Figueiredo thanks J. Bioucas-Dias, R. Nowak, and S. Wright, with whom he worked on image denoising and deblurring. Y. Ma thanks his former students and colleagues on the work of generalized PCA and image superresolution: H. Derksen, R. Fossum, W. Hong, K. Huang, T. Huang, S. Sastry, R. Vidal, J. Wright, and J. Yang.

REFERENCES

- [1] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, pp. 4311–4322, Nov. 2006.
- [2] H. Andrews and B. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.
- [3] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
- [4] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A ℓ^1 -unified variational framework for image restoration," in *Proc. Eur. Conf. Comput. Vision (ECCV2004)*, 2004, pp. 1–13.
- [5] J. Bioucas-Dias and M. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, pp. 2992–3004, 2007.
- [6] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

- [7] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modelling of signals and images," *SIAM Rev.*, vol. 51, pp. 34–81, 2009.
- [8] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR'05)*, San Diego, CA, 2005, pp. 60–65.
- [9] E. J. Candès and D. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities," *Commun. Pure Appl. Math.*, vol. 57, pp. 219–266, 2004.
- [10] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n ," *Ann. Statistics.*, vol. 35, pp. 2313–2351, 2005.
- [11] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 24, pp. 21–30, Mar. 2008.
- [12] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *Handbook of Mathematical Models in Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. Berlin, Germany: Springer-Verlag, 2005.
- [13] S. Chen, D. Donoho, and M. Saunders, "Atomic decompositions by basis pursuit," *SIAM Rev.*, vol. 43, pp. 129–159, 2001.
- [14] R. Coifman and D. L. Donoho, "Translation invariant de-noising," in *Wavelets and Statistics*, vol. 103, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995, pp. 125–150.
- [15] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM J. Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.
- [16] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, "Soft edge smoothness prior for alpha channel super resolution," in *Proc. Int. Conf. Comput. Vision*, 2007.
- [17] I. Daubechies, M. Defriese, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. LVII, pp. 1413–1457, 2004.
- [18] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, pp. 613–627, 1995.
- [19] D. L. Donoho and I. Johnstone, "Ideal adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [20] D. L. Donoho and J. Tanner, "Neighborliness of randomly-projected simplices in high dimensions," *Proc. Nat. Acad. Sci.*, vol. 102, pp. 9452–9457, 2005.
- [21] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *Proc. Int. Conf. Computer Vision (ICCV'99)*, Corfu, Greece, 1999, pp. 1033–1038.
- [22] M. Elad, B. Matalon, and M. Zibulevsky, "Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization," *Appl. Comput. Harmon. Anal.*, vol. 23, pp. 346–367, 2007.
- [23] M. Elad, "Why simple shrinkage is still relevant for redundant representations?" *IEEE Trans. Inf. Theory*, vol. 52, pp. 5559–5569, 2006.
- [24] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, pp. 3736–3745, Dec. 2006.
- [25] M. Elad, B. Matalon, and M. Zibulevsky, "Image denoising with shrinkage and redundant representations," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit. (CVPR'2006)*, New York, 2006.
- [26] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Prob.*, vol. 23, pp. 947–968, 2007.
- [27] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *J. Appl. Comput. Harmon. Anal.*, vol. 19, pp. 340–358, Nov. 2005.
- [28] M. Figueiredo and R. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffreys' noninformative prior," *IEEE Trans. Image Process.*, vol. 10, pp. 1322–1331, 2001.
- [29] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, 2003.
- [30] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Proc. IEEE Int. Conf. Image Process. (ICIP'05)*, Genoa, Italy, 2005.
- [31] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, 2002.
- [32] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, 1984.
- [33] O. G. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising—Part I: Theory," *IEEE Trans. Image Process.*, vol. 15, pp. 539–554, Mar. 2006.
- [34] O. G. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising—Part II: Adaptive algorithms," *IEEE Trans. Image Process.*, vol. 15, pp. 555–571, Mar. 2006.
- [35] W. Hong, J. Wright, K. Huang, and Y. Ma, "Multi-scale hybrid linear models for lossy image representation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3655–3671, Dec. 2006.
- [36] A. Hyvärinen, "Sparse code shrinkage: Denoising of non-Gaussian data by maximum likelihood estimation," *Neural Comput.*, vol. 11, pp. 1739–1768, 1999.
- [37] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Process. Lett.*, vol. 3, pp. 10–12, 1996.
- [38] Y. Ma, A. Yang, H. Derksen, and R. Fofsum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Rev.*, vol. 50, no. 3, Aug. 2008.
- [39] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy coding and compression," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, pp. 1546–1562, Sep. 2007.
- [40] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, pp. 53–69, Jan. 2008.
- [41] S. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, 1989.
- [42] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1998.
- [43] M. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, pp. 300–303, 1999.
- [44] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, pp. 909–919, 1999.
- [45] D. Mumford, "Empirical statistics and stochastic models for visual signals," in *New Directions in Statistical Signal Processing: From Systems to Brain*, S. Haykin, J. Principe, T. Sejnowski, and J. McWhirter, Eds. Cambridge, MA: MIT Press, 2005.
- [46] R. Nowak and M. Figueiredo, "Fast wavelet-based image deconvolution using the EM algorithm," in *Proc. 35th Asilomar Conf. Signals, Syst., Comput.*, 2001, vol. 1, pp. 371–375.
- [47] B. Olshausen and D. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.
- [48] G. Peyré, "Manifold models for signals and images," *Comput. Vision Image Understand.*, vol. 113, pp. 249–260, 2009.
- [49] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.
- [50] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inf. Theory*, vol. 38, pp. 587–607, Mar. 1992.
- [51] E. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Int. Conf. Image Process.*, Lausanne, Switzerland, 1996, pp. 379–382.
- [52] J.-L. Starck, E. Candès, and D. Donoho, "Astronomical image representation by the curvelet transform," *Astron. Astrophys.*, vol. 398, pp. 785–800, 2003.
- [53] J.-L. Starck, M. Nguyen, and F. Murtagh, "Wavelets and curvelets for image deconvolution: A combined approach," *Signal Process.*, vol. 83, pp. 2279–2283, 2003.
- [54] M. Vetterli, "A theory of multirate filter banks," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 356–372, 1987.
- [55] B. Vidakovic, "Nonlinear wavelet shrinkage with Bayes rules and Bayes factors," *J. Amer. Statist. Assoc.*, vol. 93, pp. 173–179, 1998.
- [56] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, pp. 1945–1959, Dec. 2005.
- [57] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, 2009.
- [58] J. Yang, H. Tang, Y. Ma, and T. Huang, "Face hallucination via sparse coding," in *Proc. Int. Conf. Image Process.*, 2008.
- [59] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2008.

ABOUT THE AUTHORS

Michael Elad (Senior Member, IEEE) received the B.Sc., M.Sc., and D.Sc. degrees from the Department of Electrical Engineering, The Technion—Israel Institute of Technology, Haifa, in 1986, 1988, and 1997, respectively.

From 1988 to 1993, he served in the Israeli Air Force. From 1997 to 2001, he held research and management positions in industry. From 2001 to 2003, he spent a postdoctoral period as a Research Associate with the Computer Science Department, Stanford University (SCCM Program). Since September 2003, he has been with the Department of Computer Science, The Technion. In May 2007, he was tenured as an Associate Professor. He works in the field of signal and image processing, specializing in inverse problems and sparse representations. He is an Associate Editor for the *SIAM Journal on Imaging Sciences*.

He received The Technion's Best Lecturer Award seven times. He received the Solomon Simon Mani Award for excellence in teaching (2007) and the Henri Taub Prize for academic excellence (2008). He is an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING.



Mário A. T. Figueiredo (Fellow, IEEE) received the E.E., M.Sc., Ph.D., and Agregado degrees in electrical and computer engineering from Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal, in 1985, 1990, 1994, and 2004, respectively.

Since 1994, he has been with the Faculty of the Department of Electrical and Computer Engineering, IST. He is also Area Coordinator with Instituto de Telecomunicações, a private not-for-profit research institution. His scientific interests include image processing and analysis, computer vision, statistical pattern recognition, and statistical learning.

Dr. Figueiredo is a Fellow of the International Association for Pattern Recognition. He received the 1995 Portuguese IBM Scientific Prize and the 2008 UTL/Santander-Totta Scientific Prize. He is a member of the IEEE Image, Video, and Multidimensional Signal Processing Technical Committee and has been an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON MOBILE COMPUTING, and PATTERN RECOGNITION LETTERS. He has been a Guest Coeditor of special issues of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He was a Cochair of the 2001 and 2003 Workshops on Energy Minimization Methods in Computer Vision and Pattern Recognition and Program/Technical Committee member of many international conferences.



Yi Ma (Senior Member, IEEE) received the bachelor's degree in automation and in applied mathematics from Tsinghua University, Beijing, China, in 1995. He received the M.S. degree in electrical engineering and computer science (EECS), the M.A. degree in mathematics, and the Ph.D. degree in EECS from the University of California, Berkeley, in 1997, 2000, and 2000, respectively.

Since 2000, he has been on the Faculty of the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, where he now is an Associate Professor. Currently, he is also the Research Manager for the Visual Computing Group, Microsoft Research Asia, Beijing. His main research areas are in systems theory and computer vision.

Prof. Ma is a member of ACM. He received the David Marr Best Paper Prize at the International Conference on Computer Vision in 1999 and Honorable Mention for the Longuet-Higgins Best Paper Award at the European Conference on Computer Vision in 2004. He received the CAREER Award from the National Science Foundation in 2004 and the Young Investigator Program Award from the Office of Naval Research in 2005. He is an Associate Editor for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE.

