Image Decomposition via the Combination of Sparse Representations and a Variational Approach

Jean-Luc Starck, Michael Elad, and David L. Donoho

***Abstract*—The separation of image content into semantic parts plays a vital role in applications such as compression, enhancement, restoration, and more. In recent years, several pioneering works suggested such a separation be based on variational formulation and others using independent component analysis and sparsity. This paper presents a novel method for separating images into texture and piecewise smooth (cartoon) parts, exploiting both the variational and the sparsity mechanisms. The method combines the basis pursuit denoising (BPDN) algorithm and the total-variation (TV) regularization scheme. The basic idea presented in this paper is the use of two appropriate dictionaries, one for the representation of textures and the other for the natural scene parts assumed to be piecewise smooth. Both dictionaries are chosen such that they lead to sparse representations over one type of image-content (either texture or piecewise smooth). The use of the BPDN with the two amalgamed dictionaries leads to the desired separation, along with noise removal as a by-product. As the need to choose proper dictionaries is generally hard, a TV regularization is employed to better direct the separation process and reduce ringing artifacts. We present a highly efficient numerical scheme to solve the combined optimization problem posed by our model and to show several experimental results that validate the algorithm's performance.**

***Index Terms*—Basis pursuit denoising (BPDN), curvelet, local discrete cosine transform (DCT), piecewise smooth, ridgelet, sparse representations, texture, total variation, wavelet.**

I. Introduction

THE TASK of decomposing signals into their building atoms is of great interest in many applications. The typical assumption made in such problems is that the given signal is a linear mixture of several source signals of a more coherent origin. These kinds of problems have drawn a lot of research attention recently. Independent component analysis (ICA), sparsity methods, and variational calculus, have all been used for the separation of signal mixtures with varying degrees of success (see, for example, [1]–[5]). A classic example is the cocktail party problem where a sound signal containing several concurrent speakers is to be decomposed into the separate speakers. In image processing, a parallel situation is encountered in cases of photographs containing transparent layers due to reflection.

Manuscript received February 18, 2004; revised August 23, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Truong Q. Nguyen.

J.-L. Starck is with the CEA-Saclay, DAPNIA/SEDI-SAP, Service d'Astrophysique, F-91191 Gif sur Yvette, France (e-mail: jstarck@cea.fr).

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

D. L. Donoho is with the Department of Statistics, Stanford University, Stanford, CA 94305 USA (e-mail: donoho@stat.stanford.edu).

Digital Object Identifier 10.1109/TIP.2005.852206

An interesting decomposition application—separating texture from nontexture parts in images—has been recently studied by several researchers. The importance of such separation is for applications in image compression, image analysis, synthesis and more (see, for example, [6]). A variational-based method was proposed recently by Vese and Osher [3] and later followed by others [5], [7], [8]. Their approach uses a recently introduced mathematical model for texture content [9] that extends the notion of total-variation (TV) [10]. A different methodology toward the same separation task is proposed in [2] and [4]. The work in [2] describes a novel image compression algorithm based on image decomposition to cartoon and texture layers using the wavelet-packet transform. The work presented in [4] shows a separation based on the matching pursuit algorithm and an MRF modeling. We will return to these works and give a more detailed description of their contribution and their relation to the work presented here.

In this paper, we focus on the same decomposition problem—texture and natural (piecewise smooth) additive ingredients. Fig. 1 presents the desired behavior of the separation task at hand for a typical example. In this work, we aim at separating these two parts on a pixel-by-pixel basis, such that if the texture appears in parts of the spatial support of the image, the separation should succeed in finding a masking map as a by-product of the separation process.

The approach we take for achieving the separation starts with the basis pursuit denoising (BPDN) algorithm, extending results from previous work [11], [12]. The core idea here is to choose two appropriate dictionaries, one for the representation of texture, and the other for the natural scene parts. Both dictionaries are to be chosen such that each leads to sparse representations over the images it is serving, while yielding nonsparse representations on the other content type. Thus, when amalgamated to one dictionary, the BPDN is expected to lead to the proper separation, as it seeks for the overall sparsest solution, and this should align with the sparse representation for each part. We show experimentally how indeed the BPDN framework leads to a successful separation. Furthermore, we show how to strengthen the BPDN paradigm, overcoming ringing artifacts by leaning on the TV regularization scheme.

The rest of the paper is organized as follows. Section II presents the separation method, how the BPDN is used, and how TV is added to obtain a further improvement. In Section III, we discuss the choice of the dictionaries for the texture and the natural scene parts. Section IV addresses the numerical scheme for solving the separation problem efficiently. We present several experimental results in Section V. Relation to prior art relevant to this work is presented in Section VI, and

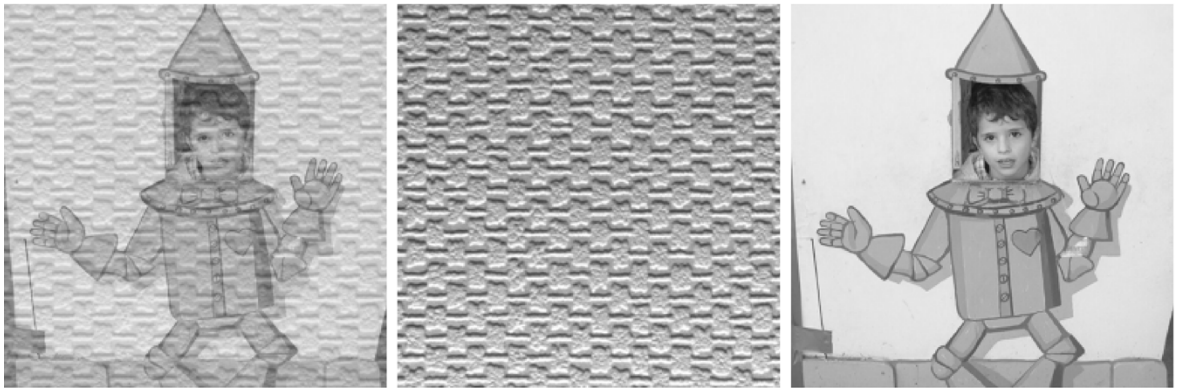

Fig. 1. Example of a separation of texture from piecewise smooth content in an image.

conclusions are drawn in Section VII. Two appendices in this paper give a detailed presentation of a numerical algorithm that is found useful here, and a preliminary theoretical study of the separation task.

II. Separation of Images—Basics

A. Model Assumption

Assume that the input image to be processed is of size $N \times N$. We represent this image as a one-dimensional (1-D) vector of length N^2 by simple reordering. For such images $\underline{X}_t$ that contain *only* pure texture content we propose an over-complete representation matrix $\mathbf{T}_t \in \mathcal{M}^{N^2 \times L}$ (where typically $L \gg N^2$) such that solving

$$\underline{\alpha}_t^{\text{opt}} = \text{Arg}\min_{\underline{\alpha}_t} \|\underline{\alpha}_t\|_0 \text{ subject to : } \underline{X}_t = \mathbf{T}_t\underline{\alpha}_t \qquad (1)$$

for any texture image $\underline{X}_t$ leads to a very sparse solution. The notation $\|\underline{u}\|_0$ is the ℓ^0 norm of the vector $\underline{u}$, effectively counting the number of nonzeros in it. We further assume that $\mathbf{T}_t$ is such that if the texture appears in parts of the image and otherwise zero, the representation is still sparse, implying that the dictionary employs a multiscale and local analysis of the image content. The definition in (1) is essentially an overcomplete transform of $\underline{X}_t$, yielding a representation $\underline{\alpha}_t$, such that sparsity is maximized.

We further require that when this forward transform with $\mathbf{T}_t$ is applied to images containing no texture and pure piecewise-smooth content, the resulting representations are nonsparse. Thus, the dictionary $\mathbf{T}_t$ plays a role of a discriminant between content types, preferring the texture over the natural part. A possible measure of fidelity of the chosen dictionary is the functional

$$\begin{aligned}
\mathbf{T}_t^{\text{opt}} &= \text{Arg}\min_{\mathbf{T}_t} \frac{\sum_k \|\underline{\alpha}_t^{\text{opt}}(k)\|_0}{\sum_j \|\underline{\alpha}_n^{\text{opt}}(j)\|_0} \\
&\text{where : } \underline{\alpha}_t^{\text{opt}}(k) = \text{Arg}\min_{\underline{\alpha}_t} \|\underline{\alpha}_t\|_0 \\
&\qquad \text{subject to : } \underline{X}_t(k) = \mathbf{T}_t\underline{\alpha}_t, \ \ k = 1, 2, \ \ldots \\
&\qquad \underline{\alpha}_n^{\text{opt}}(j) = \text{Arg}\min_{\underline{\alpha}_n} \|\underline{\alpha}_n\|_0 \\
&\qquad \text{subject to : } \underline{X}_n(j) = \mathbf{T}_t\underline{\alpha}_n, \ \ j = 1, 2, \ \ldots . \qquad (2)
\end{aligned}$$

This functional of the dictionary is measuring the relative sparsity between a family of textured images $\{\underline{X}_t(k)\}_k$ and a family of natural content images $\{\underline{X}_n(j)\}_j$. This, or a similar measure, could be used for the design of the proper choice of $\mathbf{T}_t$. However, in this paper, we take a different approach, as will be discussed shortly.

Similar to the above, assume that for images containing piecewise smooth content $\underline{X}_n$, we have a different dictionary $\mathbf{T}_n$, such that their content is sparsely represented by the above definition. Again, we assume that beyond the sparsity obtained by $\mathbf{T}_n$ for natural images, we can further assume that texture images are represented very inefficiently (i.e., nonsparsely) and also assume that the analysis applied by this dictionary is of multiscale and local nature, enabling it to detect pieces of the desired content.

For an arbitrary image $\underline{X}$ containing both texture and piecewise smooth content (overlayed, side-by-side, or both), we propose to seek the sparsest of all representations over the augmented dictionary containing both $\mathbf{T}_t$ and $\mathbf{T}_n$. Thus, we need to solve

$$\begin{aligned}
\{\underline{\alpha}_t^{\text{opt}}, \underline{\alpha}_n^{\text{opt}}\} &= \text{Arg}\min_{\{\underline{\alpha}_t, \underline{\alpha}_n\}} \|\underline{\alpha}_t\|_0 + \|\underline{\alpha}_n\|_0 \\
&\text{subject to : } \underline{X} = \mathbf{T}_t\underline{\alpha}_t + \mathbf{T}_n\underline{\alpha}_n. \qquad (3)
\end{aligned}$$

This optimization task is likely to lead to a successful separation of the image content, such that $\mathbf{T}_t\underline{\alpha}_t$ is mostly texture and $\mathbf{T}_n\underline{\alpha}_n$ is mostly piecewise smooth. This expectation relies on the assumptions made earlier about $\mathbf{T}_t$ and $\mathbf{T}_n$ being very efficient in representing one content type and being highly ineffective in representing the other.

While sensible from the point of view of the desired solution, the problem formulated in (3) is nonconvex and hard to solve. Its complexity grows exponentially with the number of columns in the overall dictionary. The basis pursuit (BP) method [11] suggests the replacement of the ℓ^0 norm with an ℓ^1 norm, thus leading to a solvable optimization problem (linear programming) of the form

$$\begin{aligned}
\{\underline{\alpha}_t^{\text{opt}}, \underline{\alpha}_n^{\text{opt}}\} &= \text{Arg}\min_{\{\underline{\alpha}_t, \underline{\alpha}_n\}} \|\underline{\alpha}_t\|_1 + \|\underline{\alpha}_n\|_1 \\
&\text{subject to : } \underline{X} = \mathbf{T}_t\underline{\alpha}_t + \mathbf{T}_n\underline{\alpha}_n. \qquad (4)
\end{aligned}$$

Interestingly, recent work have shown that for sparse enough solutions, the BP simpler form is accurate, also leading to the

sparsest of all representations [13]–[16]. More about this relationship is given in Appendix II, where we analyze theoretically bounds on the success of such separation.

B. Complicating Factors

The above description is sensitive in a way that may hinder the success of the overall separation process. There are two complicating factors, both have to do with the assumptions made above.

Assumption: The image is decomposed cleanly into texture and natural (piecewise smooth) parts. For an arbitrary image, this assumption is not true, as it may also contain additive noise that is not represented well both by $\mathbf{T}_t$ and $\mathbf{T}_n$. Generally speaking, any deviation from this assumption may lead to a nonsparse pair of vectors $\{\underline{\alpha}_t^{\text{opt}}, \underline{\alpha}_n^{\text{opt}}\}$, and with that, due to the change from ℓ^0 to ℓ^1, to a complete failure of the separation process.

Assumption: The chosen dictionaries are appropriate. It is very hard to propose a dictionary that leads to sparse representations for a wide family of signals. A chosen dictionary may be inappropriate because it does not lead to a sparse representation for the proper signals. If this is the case, then, for such images, the separation will fail. A more complicating scenario is obtained for dictionaries that does not discriminate well between the two phenomena we desire to separate. Thus, if, for example, we have a dictionary $\mathbf{T}_n$ that indeed leads to sparse representations for natural scenes, but also known to lead to sparse representations for some texture content, clearly, such a dictionary could not be used for a successful separation. Put more generally, we may ask whether such dictionaries exist at all.

A solution for the first problem could be obtained by relaxing the constraint in (4) to become an approximate one. Thus, in the new form, we propose the solution of

$$\{\underline{\alpha}_t^{\text{opt}}, \underline{\alpha}_n^{\text{opt}}\} = \text{Arg}\min_{\{\underline{\alpha}_t, \underline{\alpha}_n\}} \|\underline{\alpha}_t\|_1 + \|\underline{\alpha}_n\|_1 + \lambda\|\underline{X} - \mathbf{T}_t\underline{\alpha}_t - \mathbf{T}_n\underline{\alpha}_n\|_2^2. \quad (5)$$

Thus, an additional content in the image that is not represented sparsely by both dictionaries will be allocated to be the residual $\underline{X} - \mathbf{T}_t\underline{\alpha}_t - \mathbf{T}_n\underline{\alpha}_n$. This way, we not only manage to separate texture from natural scene parts, but also succeed in removing an additive noise as a by-product. This new formulation is familiar by the name BP denoising, shown in [11] to perform well for denoising tasks. We should note here that the choice of ℓ^2 as the error norm is intimately related to the assumption that the residual behaves like a white zero-mean Gaussian noise. Other norms can be similarly introduced to account for different noise models, such as Laplacian (ℓ^1), uniformly distributed noise (ℓ^∞), and others.

As for the second problem mentioned here, we propose an underlying model to describe image content, but we do not, and cannot, claim that this model is universal and will apply to all images. There are certainly images for which this model will fail. Still, in properly choosing the dictionaries, the proposed model holds true for a relatively large class of images. Indeed, the experimental results to follow support this belief.

Also, even if the above-described model is feasible, the problem of choosing the proper dictionaries remains open and difficult. This matter will be discussed in the next section.

Suppose we have chosen $\mathbf{T}_n$ and $\mathbf{T}_t$, both generally well suited for the separation task. By adding external forces that direct the images $\mathbf{T}_n\underline{\alpha}_n$ and $\mathbf{T}_t\underline{\alpha}_t$ to better suite their expected content, these forces will fine tune the process to achieve its task. As an example for such successful external force, adding a TV penalty [10] to (5) can direct the image $\mathbf{T}_n\underline{\alpha}_n$ to fit the piecewise smooth model. This leads to

$$\{\underline{\alpha}_t^{\text{opt}}, \underline{\alpha}_n^{\text{opt}}\} = \text{Arg}\min_{\{\underline{\alpha}_t, \underline{\alpha}_n\}} \|\underline{\alpha}_t\|_1 + \|\underline{\alpha}_n\|_1 + \lambda\|\underline{X} - \mathbf{T}_t\underline{\alpha}_t - \mathbf{T}_n\underline{\alpha}_n\|_2^2 + \gamma TV\{\mathbf{T}_n\underline{\alpha}_n\}. \quad (6)$$

The expression $TV\{\mathbf{T}_n\underline{\alpha}_n\}$ is essentially computing the image $\underline{X}_n = \mathbf{T}_n\underline{\alpha}_n$ (supposed to be piecewise smooth) and applying the TV norm on it (computing its absolute gradient field and summing it with an ℓ^1 norm). Penalizing with TV, we force the image $\mathbf{T}_n\underline{\alpha}_n$ to be closer to a piecewise smooth image and, thus, support the separation process. This idea has already appeared in [17]–[19], where TV was used to damp ringing artifacts near edges, caused by the oscillations of the curvelet atoms. We note that combining TV with wavelet has also been done for similar reasons in [20], although in a different fashion.

C. Different Problem Formulation

Assume that each of the chosen dictionaries can be composed into a set of unitary matrices such that

$$\mathbf{T}_t = [\mathbf{T}(1)_t, \mathbf{T}(2)_t, \ldots, \mathbf{T}(L_t)_t]$$
$$\mathbf{T}_n = [\mathbf{T}(1)_n, \mathbf{T}(2)_n, \ldots, \mathbf{T}(L_n)_n]$$

and

$$\begin{aligned}\mathbf{T}(1)_t^H\mathbf{T}(1)_t &= \mathbf{T}(2)_t^H\mathbf{T}(2)_t = \cdots = \mathbf{T}(L_t)_t^H\mathbf{T}(L_t)_t \\ &= \mathbf{T}(1)_n^H\mathbf{T}(1)_n = \mathbf{T}(2)_n^H\mathbf{T}(2)_n \\ &= \cdots = \mathbf{T}(L_n)_n^H\mathbf{T}(L_n)_n = \mathbf{I}\end{aligned}$$

where $\mathbf{T}^H$ is the Hermite adjoint (conjugate and transpose) of $\mathbf{T}$. In such a case we could slice $\underline{\alpha}_t$ and $\underline{\alpha}_n$ into L_t and L_n parts, correspondingly, and obtain a new formulation of the problem

$$\begin{aligned}\min_{\{\underline{\alpha}(k)_t\}_{k=1}^{L_t}, \{\underline{\alpha}(j)_n\}_{j=1}^{L_n}} & \sum_{k=1}^{L_t}\|\underline{\alpha}(k)_t\|_1 + \sum_{j=1}^{L_n}\|\underline{\alpha}(j)_n\|_1 \\ &+ \lambda\left\|\underline{X} - \sum_{k=1}^{L_t}\mathbf{T}(k)_t\underline{\alpha}(k)_t - \sum_{j=1}^{L_n}\mathbf{T}(j)_n\underline{\alpha}(j)_n\right\|_2^2 \\ &+ \gamma TV\left\{\sum_{j=1}^{L_n}\mathbf{T}(j)_n\underline{\alpha}(j)_n\right\}. \end{aligned} \quad (7)$$

In the above formulation, the representation vector pieces $\underline{\alpha}(j)_n$ and $\underline{\alpha}(k)_t$ are supposed to be sparse. Defining $\underline{X}(k)_t = \mathbf{T}(k)_t\underline{\alpha}(k)_t$ and similarly $\underline{X}(j)_n = \mathbf{T}(j)_n\underline{\alpha}(j)_n$, we can reformulate the problem as

$$\min_{\{\underline{X}(k)_t\}_{k=1}^{L_t}, \{\underline{X}(j)_n\}_{j=1}^{L_n}} \sum_{k=1}^{L_t}\left\|\mathbf{T}(k)_t^H\underline{X}(k)_t\right\|_1$$

$$+\sum_{j=1}^{L_n}\left\|\mathbf{T}(j)_n^H \underline{X}(j)_n\right\|_1 + \lambda\left\|\underline{X}-\sum_{k=1}^{L_t}\underline{X}(k)_t-\sum_{j=1}^{L_n}\underline{X}(j)_n\right\|_2^2+\gamma TV\left\{\sum_{j=1}^{L_n}\underline{X}(j)_n\right\} \quad (8)$$

and the unknowns become images, rather then representation coefficients. For this problem structure, there exists a fast numerical solver called the *block-coordinate relaxation method*, based on the shrinkage method [21]. This solver (see Appendix I for details) requires *only* the use of matrix-vector multiplications with the unitary transforms and their adjoints. See [22] for more details. We will return to this form of solution when we discuss numerical algorithms.

D. Summary of Method

In order to translate the above ideas into a practical algorithm, we should answer three major questions: 1) Is there a theoretical backup to the heuristic claims made here? 2) How should we choose the dictionaries $\mathbf{T}_t$ and $\mathbf{T}_n$? 3) How should we numerically solve the obtained optimization problem in a traceable way? These three questions are addressed in the coming sections. The theoretical grounds for the separation is briefly discussed in Appendix II. The choice of dictionaries in the topic of the next section, and the numerical considerations follow in Section IV.

III. Candidate Dictionaries

Our approach toward the choice of $\mathbf{T}_t$ and $\mathbf{T}_n$ is to pick known transforms and not design those optimally as we hinted earlier as a possible method. We choose transforms known for representing well either texture or piecewise smooth behaviors. For numerical reasons, we restrict our choices to the dictionaries $\mathbf{T}_t$ and $\mathbf{T}_n$ that have a fast forward and inverse implementation. In making a choice for a transform, we use experience of the user applying the separation algorithm, and, thus, the choices made may vary from one image to another. We shall start with a brief description of our candidate dictionaries.

A. Dictionaries for Piecewise Smooth Content

1) Bi-Orthogonal Wavelet Transforms (OWT): Previous work has established that the wavelet transform is well suited for the effective (sparse) representation of natural scene [21]. The application of the OWT to image compression using the 7–9 filters and the zero-tree coding leads to impressive results over the JPEG [23]–[25].

The OWT implementation requires $O(N^2)$ operations for an image with $N \times N$ pixels, both for the forward and the inverse transforms. Represented as a matrix-vector multiplication, this transform is a square matrix, either unitary, or nonunitary with accompanying inverse matrix of a similar simple form. The OWT presents only a fixed number of directional elements independent of scales, and there is no highly anisotropic elements [26]. Therefore, we expect the OWT to be nonoptimal for detection of highly anisotropic features. Moreover, the OWT is nonshift invariance—a property that may cause difficulties in our analysis.

The undecimated version (UWT) of the OWT is certainly the most popular transform for data filtering. It is obtained by skipping the decimation, implying that this is an overcomplete transform represented as a matrix with more columns than rows. The redundancy factor (ratio between number of columns to number of rows) is $3J+1$, where J is the number of resolution layers. With this over-completeness, we obtain the desired shift invariance property.

2) Isotropic à Trous Algorithm: This transform decomposes an $N \times N$ image I as a superposition of the form $I(x,y) = c_J(x,y) + \sum_{j=1}^{J} w_j(x,y)$, where c_J is a coarse or smooth version of the original image I and w_j represents the details of I at scale 2^{-j} (see [27]). Thus, the algorithm outputs $J+1$ subband arrays of size $N \times N$. This wavelet transform is very well adapted to the detection of isotropic features, and this explains the reason of its success for astronomical image processing, where the data contains mostly (quasi-)isotropic objects, such as stars or galaxies [28].

3) Local Ridgelet Transform: The ridgelet transform is the application of a 1-D wavelet to the angular slices of the Radon transform [26]. Such transform has been shown to be very effective for representing global lines in an image. In order to detect line segments, a partitioning must be introduced [29], and a ridgelet transform is to be applied per each block. In such a case, the image is decomposed into 50% overlapping blocks of side-length b pixels. The overlap is introduced in order to avoid blocking artifacts. For a $N \times N$ image, we count $2N/b$ such blocks in each direction. The overlap introduces more redundancy (over-completeness), as each pixel belongs to four neighboring blocks. The ridgelet transform requires $O(N^2 \log_2 N)$ operations. More details on the implementation of the digital ridgelet transform can be found in [30].

4) Curvelet Transform: The curvelet transform, proposed in [31], [32], and [30], enables the directional analysis of an image in different scales. The idea is to first decompose the image into a set of wavelet bands, and to analyze each band with a local ridgelet transform. The block size is changed at each scale level, such that different levels of the multiscale ridgelet pyramid are used to represent different subbands of a filter bank output. The side-length of the localizing windows is doubled *at every other* dyadic subband, hence maintaining the fundamental property of the curvelet transform, which says that elements of length about $2^{-j/2}$ serve for the analysis and synthesis of the jth subband $[2^j, 2^{j+1}]$. The curvelet transform is also redundant, with a redundancy factor of $16J+1$ whenever J scales are employed. Its complexity is of the $O(N^2 \log_2 N)$, as in ridgelet. This method is best for the detection of anisotropic structures and smooth curves and edges of different lengths.

B. Dictionaries for Texture Content

1) (Local) Discrete Cosine Transform (DCT): The DCT is a variant of the Discrete Fourier Transform, replacing the complex analysis with real numbers by a symmetric signal extension. The DCT is an orthonormal transform, known to be well suited for first order Markov stationary signals. Its coefficients essentially represents frequency content, similar

to the ones obtained by Fourier analysis. When dealing with nonstationary sources, DCT is typically applied in blocks. Such is indeed the case in the JPEG image compression algorithm. Choice of overlapping blocks is preferred for analyzing signals while preventing artefact. In such a case we get again an overcomplete transform with redundancy factor of 4 for an overlap of 50%. A fast algorithm with complexity of $N^2 \log_2 N$ exists for its computation. The DCT is appropriate for a sparse representation of either smooth or periodic behaviors.

2) Gabor Transform: The Gabor transform is quite popular among researchers working on texture content. This transform is essentially a localized DFT, where the localization is obtained by windowing portions of the signal in an overlapping fashion. The amount of redundancy is controllable. For a proper choice of the overlap and the window, both the forward and the inverse transforms can be applied with complexity of $N^2 \log_2 N$.

IV. Numerical Considerations

A. Numerical Scheme

Returning to the separation process as posed in (6), we need to solve an optimization problem of the form

$$\{\underline{\alpha}_t^{\text{opt}}, \underline{\alpha}_n^{\text{opt}}\} = \text{Arg} \min_{\{\underline{\alpha}_t, \underline{\alpha}_n\}} \|\underline{\alpha}_t\|_1 + \|\underline{\alpha}_n\|_1 + \lambda \|\underline{X} - \mathbf{T}_t\underline{\alpha}_t - \mathbf{T}_n\underline{\alpha}_n\|_2^2 + \gamma TV\{\mathbf{T}_n\underline{\alpha}_n\}. \quad (9)$$

Instead of solving this optimization problem, finding $\{\underline{\alpha}_t^{\text{opt}}, \underline{\alpha}_n^{\text{opt}}\}$, let us reformulate the problem so as to get the texture and the natural part images, $\underline{X}_t$ and $\underline{X}_n$, as our unknowns. The reason behind this change is the obvious simplicity gained by searching lower dimensional vectors—representation vectors are far longer than the images they represent for overcomplete dictionaries as the ones we use here.

Define $\underline{X}_t = \mathbf{T}_t\underline{\alpha}_t$ and $\underline{X}_n = \mathbf{T}_n\underline{\alpha}_n$. Given $\underline{X}_t$, we can recover $\underline{\alpha}_t$ as $\underline{\alpha}_t = \mathbf{T}_t^+\underline{X}_t + \underline{r}_t$ where $\underline{r}_t$ is an arbitrary vector in the null-space of $\mathbf{T}_t$, and $\mathbf{T}_t^+$ is the Moore–Penrose pseudo-inverse of $\mathbf{T}_t$. Note that, for tight frames, this matrix is the same (up to a constant) as the Hermite adjoint one, and, thus, its computation is relatively easy. Put these back into (6), and, thus, we obtain

$$\begin{aligned}\{\underline{X}_t^{\text{opt}}, \underline{X}_n^{\text{opt}}\} &= \text{Arg} \min_{\{\underline{X}_t, \underline{X}_n, \underline{r}_t, \underline{r}_n\}} \|\mathbf{T}_t^+\underline{X}_t + \underline{r}_t\|_1 + \|\mathbf{T}_n^+\underline{X}_n + \underline{r}_n\|_1 \\ &\quad + \lambda \|\underline{X} - \underline{X}_t - \underline{X}_n\|_2^2 + \gamma TV\{\underline{X}_n\} \\ &\quad \text{subject to : } \mathbf{T}_t\underline{r}_t = 0\ ,\ \mathbf{T}_n\underline{r}_n = 0. \end{aligned} \quad (10)$$

The term $\mathbf{T}_t^+\underline{X}_t$ is an overcomplete linear transform of the image $\underline{X}_t$. Similarly, $\mathbf{T}_n^+\underline{X}_n$ is an overcomplete linear transform of the natural part. In our attempt to replace the representation vectors as unknowns, we see that we have a pair of residual vectors to be found as well. If we choose (rather arbitrarily at this stage) to assign those vectors as zeros we obtain the problem

$$\{\underline{X}_t^{\text{opt}}, \underline{X}_n^{\text{opt}}\} = \text{Arg} \min_{\{\underline{X}_t, \underline{X}_n\}} \|\mathbf{T}_t^+\underline{X}_t\|_1 + \|\mathbf{T}_n^+\underline{X}_n\|_1 + \lambda \|\underline{X} - \underline{X}_t - \underline{X}_n\|_2^2 + \gamma TV\{\underline{X}_n\}. \quad (11)$$

We can justify the choice $\underline{r}_t = \underline{0}$ and $\underline{r}_n = \underline{0}$ in several ways.

Bounding Function: Since (11) is obtained from (10) by choosing $\underline{r}_t = \underline{0}$, $\underline{r}_n = \underline{0}$, we necessarily get that the value of (10) (after optimization) is upper bounded by the value of (11). Thus, in minimizing (11), instead, we guarantee that the true function to be minimized is of even lower value.

Relation to the Block-Coordinate-Relaxation Algorithm: Comparing (11) to the case discussed in (8), we see a close resemblance. If we assume that the dictionaries involved are unitary, we get a complete equivalence between solving (10) and (11). In a way, we may refer to the approximation we have made here as a method to generalize the block-coordinate-relaxation method for the nonunitary case.

Relation to MAP: The expression written as a penalty function in (11) has a maximum *a posteriori* estimation flavor to it. It suggests that the given image $\underline{X}$ is known to originate from a linear combination of the form $\underline{X}_t + \underline{X}_n$, contaminated by Gaussian noise—this part comes from the likelihood function $\|\underline{X} - \underline{X}_t - \underline{X}_n\|_2^2$. For the texture image part, there is the assumption that it comes from a Gibbs distribution of the form $\text{Const} \cdot \exp(-\beta_t\|\mathbf{T}_t^+\underline{X}_t\|_1)$. As for the natural part, there is a similar assumption about the existence of a prior of the form $\text{Const} \cdot \exp(-\beta_n\|\mathbf{T}_n^+\underline{X}_n\|_1 - \gamma_n TV\{\underline{X}_n\})$. While different from our original point of view, these assumptions are reasonable and not far from the BP approach.

The bottom line to all this discussion is that we have chosen an approximation to our true minimization task and, with it, managed to get a simplified optimization problem, for which an effective algorithm can be proposed. Our minimization task is, thus, given by equation (11). The algorithm we use is based on the block-coordinate-relaxation method [22] (see Appendix I), with some required changes due to the nonunitary transforms involved, and the additional TV term. The algorithm is given as follows.

The algorithm for minimizing **(11)**. Here $\mathbf{T}_n$ is the curvelet transform, and $\mathbf{T}_t$ is the local DCT.[1]

1. Initialize $L_{\max}$, number of iterations per layer N, and threshold $\delta = \lambda \cdot L_{\max}$.
2. Perform N times:

Part A—Update of $\underline{X}_n$ assuming $\underline{X}_t$ is fixed:

– Calculate the residual $\underline{R} = \underline{X} - \underline{X}_t - \underline{X}_n$.

– Calculate the curvelet transform of $\underline{X}_n + \underline{R}$ and obtain $\underline{\alpha}_n = \mathbf{T}_n^+(\underline{X}_n + \underline{R})$.

– Soft threshold the coefficient $\underline{\alpha}_n$ with the δ threshold and obtain $\hat{\underline{\alpha}}_n$.

– Reconstruct $\underline{X}_n$ by $\underline{X}_n = \mathbf{T}_n\hat{\underline{\alpha}}_n$.

Part B—Update of $\underline{X}_t$ assuming $\underline{X}_n$ is fixed:

– Calculate the residual $\underline{R} = \underline{X} - \underline{X}_t - \underline{X}_n$.

– Calculate the local DCT transform of $\underline{X}_t + \underline{R}$ and obtain $\underline{\alpha}_t = \mathbf{T}_t^+(\underline{X}_t + \underline{R})$.

– Soft threshold the coefficient $\underline{\alpha}_t$ with the δ threshold and obtain $\hat{\underline{\alpha}}_t$.

– Reconstruct $\underline{X}_t$ by $\underline{X}_t = \mathbf{T}_t\hat{\underline{\alpha}}_t$.

[1]If the texture is the same on the whole image, then a global DCT should be preferred.

Part C—TV Consideration:

– Apply the TV correction by $\underline{X}_n = \underline{X}_n - \mu\gamma(\partial TV\{\underline{X}_n\}/\partial\underline{X}_n)$.

– The parameter μ is chosen either by a line-search minimizing the overall penalty function, or as a fixed step-size of moderate value that guarantees convergence.

3. Update the threshold by $\delta = \delta - \lambda$.
4. If $\delta > \lambda$, return to Step 2. Else, finish.

In the above algorithm, soft threshold is used due to our formulation of the ℓ^1 sparsity penalty term. However, as we have explained earlier, the ℓ^1 expression is merely a good approximation for the desired ℓ^0 one, and, thus, replacing the soft by a hard threshold toward the end of the iterative process may lead to better results.

We chose this numerical scheme over the BP interior-point approach in [11], because it presents two major advantages. 1) We do not need to keep all the transformations in memory. This is particularly important when we use redundant transformations such as the un-decimated wavelet transform or the curvelet one. Also, 2) we can add different constraints on the components. Here we applied only the *TV* constraint on one of the components, but other constraints, such as positivity, can easily be added as well. Our method allows us to build easily a dedicated algorithm which takes into account the *a priori* knowledge we have on the solution for a specific problem.

B. TV and Undecimated Haar Transform

A link between the TV and the undecimated Haar wavelet soft thresholding has been studied in [33], arguing that in the 1-D case the TV and the undecimated single resolution Haar are equivalent. When going to two-dimensional (2-D), this relation does not hold anymore, but the two approaches share some similarities. Whereas the TV introduces translation- and rotation-invariance, the undecimated 2-D Haar presents translation- and scale-invariance (being multiscale). In light of this interpretation, we can change the part C in the algorithm as described below. This method is expected to lead to similar results to the ones obtained with the regular TV.

Alternative Stage C—Replacement of the TV by undecimated Haar.

Part C—TV Consideration:

- Apply the TV correction by using the undecimated Haar wavelet transform $\mathcal{H}$ and a soft thresholding:
  - Calculate the undecimated Haar wavelet transform of X_n and obtain $\hat{\underline{\alpha}}_h$.
  - Soft threshold the coefficient $\underline{\alpha}_h$ with the γ threshold
  - Reconstruct $\underline{X}_n$ by $\underline{X}_n = \mathcal{H}^{-1}\hat{\underline{\alpha}}_h$.
- The parameter μ is chosen as before.

C. Noise Consideration

The case of noisy data can be easily considered in our framework, and merged into the algorithm such that we get a three-way separation to texture, natural part, and additive noise—$\underline{X} = \underline{X}_t + \underline{X}_n + \underline{V}$. We can normalize both transforms $\mathbf{T}_t^+$ and $\mathbf{T}_n^+$ such that for a given noise realization $\underline{V}$ with zero-mean and a unit standard deviation, $\alpha_n = \mathbf{T}_n^+\underline{V}$ and $\alpha_t = \mathbf{T}_t^+\underline{V}$ have also a standard deviation equals to 1. Then, only the last step of the algorithm changes by replacing the stopping criterion $\delta > \lambda$ by $\delta > k\sigma$, where σ is the noise standard deviation and $k \approx 3, 4$. This ensures that coefficients with an absolute value lower than $k\sigma$ are not taken into account.

V. Experimental Results

A. Image Decomposition

We start the description of our experiments with a synthetically generated image composed of a natural scene and a texture, where we have the ground truth parts to compare against. We implemented the proposed algorithm with the curvelet transform (five resolution levels) for the natural scene part, and a global DCT transform for the texture. We used the soft thresholding Haar as a replacement to the TV, as described in previous section. The parameter γ was fixed to 2. The overall algorithm converges in a matter of 10–20 iterations. Due to the inefficient implementation of the curvelet transform, the overall run-time of this algorithm is 30 min. Recent progress made in the implementation of the curvelet is expected to reduce this run-time by more than one order of magnitude.

In this example, we got better results if the very low frequency components of the image are first subtracted from it, and then added to $\underline{X}_n$ after the separation. The reason for this is the evident overlap that exists between the two dictionaries—both consider the low-frequency content as theirs, as both can represent it efficiently. Thus, by removing this content prior to the separation we avoid separation ambiguity. Also, by returning this content later to the curvelet part, we use our expectation to see the low frequencies as belonging to the piecewise smooth image.

Fig. 2 shows the original image (addition of the texture and the natural parts), the low frequency component, the texture reconstructed component $\underline{X}_t$ and the natural scene part $\underline{X}_n$. As can be seen, the separation is reproduced rather well. Fig. 3 shows the results of a second experiment where the separation is applied to the above combined image after being contaminated by additive noise ($\sigma = 10$). We see that the presence of noise does not deteriorate the separation algorithm's performance, and the noise is separated well.

We have also applied our method to the Barbara (512×512) image. We used the curvelet transform with the five resolution levels, and overlapping DCT transform with a block size 32×32. The parameter γ has been fixed to 0.5. Here, we used the standard TV regularization implementation. Fig. 4 shows the Barbara image, the reconstructed cosine component $\underline{X}_t$ and the reconstructed curvelet component $\underline{X}_n$. Fig. 5 shows a magnified part of the face. For comparison, the separated components reconstructed by Vese–Osher approach [3] are also shown.

We note here that in general the comparison between different image separation methods should be done with respect to the application in mind. Here we consider the separation itself as

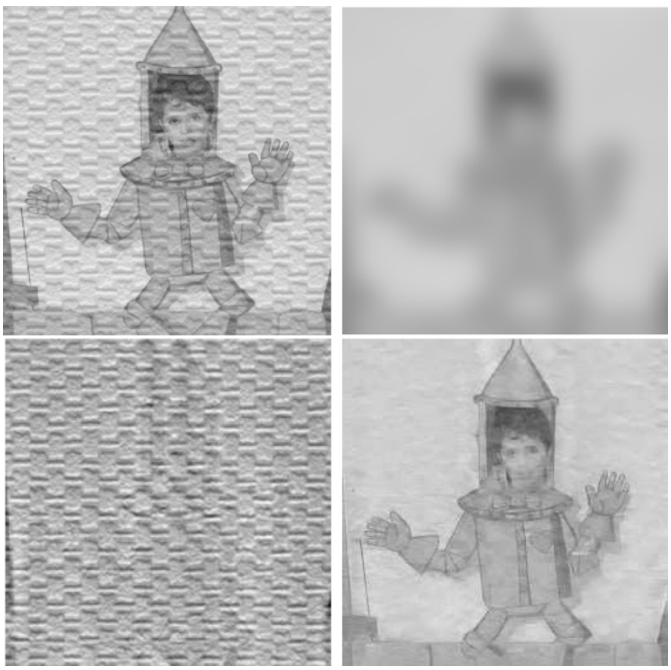

Fig. 2. (Top left) Original combined image, (top right) its low frequency content, (bottom left) the separated texture part, and (bottom right) the separated natural part.

Fig. 4. (Top) Original Barbara image. (Bottom left) The separated texture and (bottom right) the separated natural part.

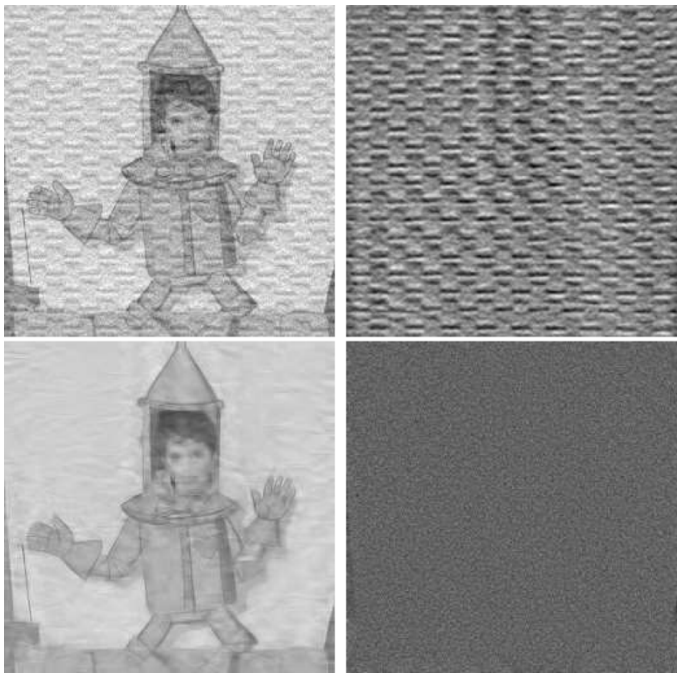

Fig. 3. (Top left) Original noisy image, (top right) the separated texture part, (bottom left) the separated natural part, (bottom right) and the residual noise component.

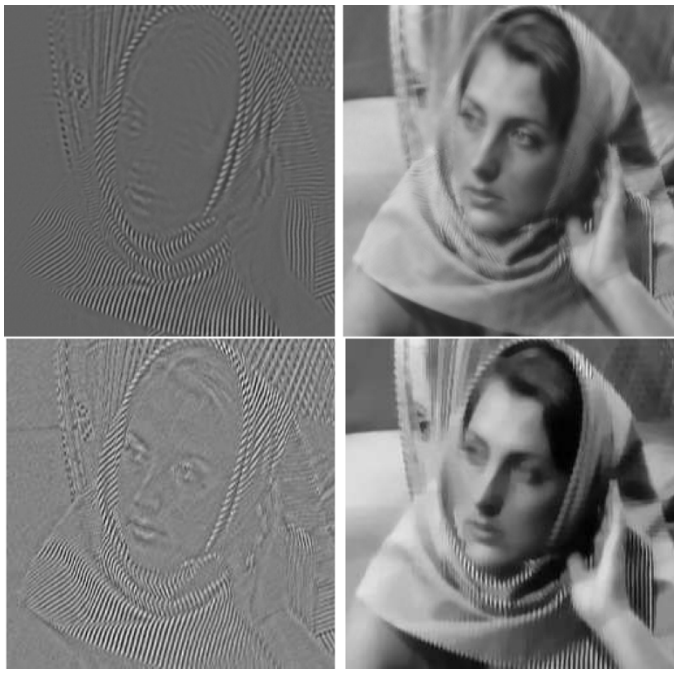

Fig. 5. Top: Reconstructed DCT and curvelet components by our method. Bottom: v and u components using Vese's algorithm.

the application, and, thus, the results are compared by visually inspecting the outcomes.

B. Nonlinear Approximation (NLA)

The efficiency of a given decomposition can be estimated by a NLA scheme, where sparsity is a measure of success. An NLA curve is obtained by reconstructing the image from the m-first best terms of the decomposition. For example, using the wavelet expansion of a function f (smooth away from a discontinuity across a C^2 curve), the best m-terms approximation $\tilde{f}_m^W$ obeys $\| f - \tilde{f}_m^W \|_2^2 \asymp m^{-1}$, $m \to \infty$, while for a Fourier expansion it is $\| f - \tilde{f}_m^F \|_2^2 \asymp m^{-1/2}$, $m \to \infty$ [34], [35]. Using the algorithm described in the previous section, we decompose the image $\underline{X}$ into two components $\underline{X}_t$ and $\underline{X}_n$ using the overcomplete transforms $\mathbf{T}_t$ and $\mathbf{T}_n$. Since the decomposition is (very)

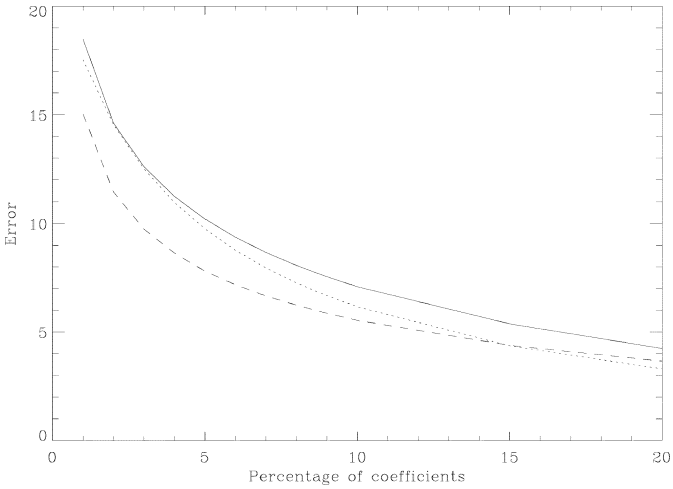

Fig. 6. Standard deviation of the error of reconstructed Barbara image versus the m largest coefficients used in the reconstruction. Full line: DCT transform. Dotted line: Orthogonal wavelet transform. Dashed line: Our signal/texture decomposition.

redundant, the exact overall representation $\underline{X}$ may require a relatively small number of coefficients due to the promoted sparsity, and this essentially yields a better NLA curve.

Fig. 6 presents the NLA curves for the image Barbara using 1) the wavelet transform (OWT), 2) the DCT, and 3) the results of the algorithm discussed here, based on the OWT-DCT combination. Denoting the wavelet transform as $\mathbf{T}_n^+$ and the DCT one as $\mathbf{T}_t^+$, the representation we use includes the m largest coefficients from $\{\underline{\alpha}_t,\ \underline{\alpha}_n\} = \{\mathbf{T}_t^+\underline{X}_t,\ \mathbf{T}_n^+\underline{X}_n\}$. Using these m values we reconstruct the image by

$$\underline{\tilde{X}}_m = \mathbf{T}_t\underline{\tilde{\alpha}}_t + \mathbf{T}_t\underline{\tilde{\alpha}}_n.$$

The curves in Fig. 6 show the representation error standard deviation as a function of m [i.e., $\mathcal{E}(m) = \sigma(\underline{X} - \underline{\tilde{X}}_m)$]. We see that for $m < 15\%$, the combined representation leads to a better NLA curve than both the DCT and the OWT alone.

C. Applications

The ability to separate the image as we show has many applications. We sketch here two such simple experiments to illustrate the importance of a successful separation.

Edge detection is a crucial processing step in many computer-vision applications. When the texture is highly contrasted, most of the detected edges are due to the texture rather than the natural part. By separating first the two components we can detect the true object's edges. Fig. 7 shows the edges detected by the Canny algorithm on both the original image and the curvelet reconstructed component (see Fig. 2).

Fig. 8 shows a galaxy imaged with the GEMINI-OSCIR instrument at 10μ. The data is contaminated by a noise and a stripping artifact (assumed to be the texture in the image) due to the instrument electronics. As the galaxy is isotropic, we used the isotropic wavelet transform instead of curvelet. Fig. 8 summarizes the results of the separation where we see a successful isolation of the galaxy, the textured disturbance, and the additive noise.

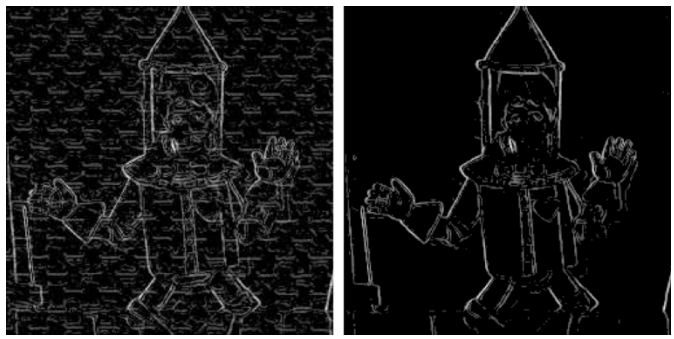

Fig. 7. Left: Detected edges on the original image. Right: Detected edges on the curvelet reconstruct component.

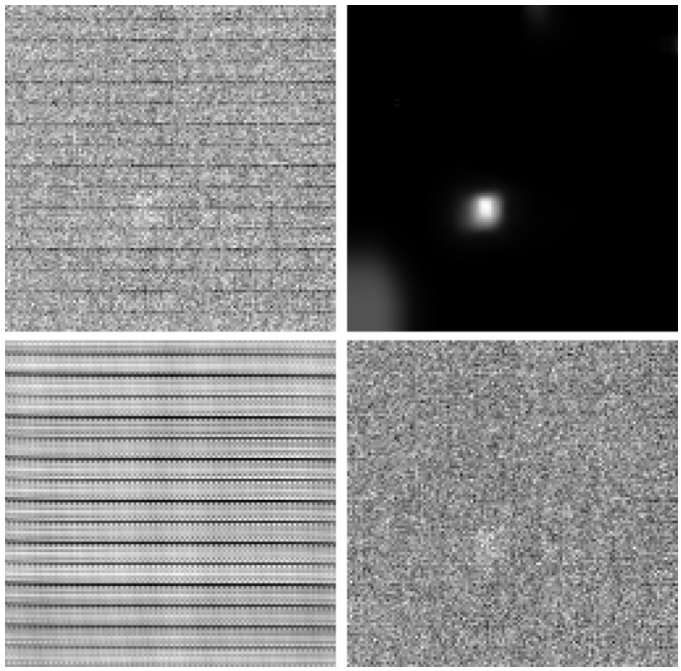

Fig. 8. (Top left) Original image. (Top right) The reconstructed wavelet component. (Bottom left) The DCT reconstructed component. (Bottom right) The residual noise.

VI. Prior Art

This work was primarily inspired by the image separation work by Vese and Osher [3]. However, there have been several other attempts to achieve such separation for various needs. We list here some of those works, present briefly their contributions, and relate them to our algorithm.

A. Variational Separation Paradigm

Whereas piecewise smooth images u are assumed to belong to the bounded-variation (*BV*) family of functions $u \in BV(\mathcal{R}^2)$, texture is known to behave differently. A different approach has recently been proposed for separating the texture v from the signal f $(= u + v)$ [3], based on a model proposed by Meyer [9]. Similar attempts and additional contributions in this line are reported in [7], [8], [36]. This model suggests that a texture image v is to belong to a different family of functions denoted as $v \in BV^*(\mathcal{R}^2)$. This notation implies the existence of two functions

$g_1, g_2 \in L^\infty(\mathcal{R}^2)$ such that $v(x,y) = \partial_x g_1(x,y) + \partial_y g_2(x,y)$. The BV^* norm is defined using the two functions g_1, g_2 as $\|v\|_{BV^*} = \|(|g_1(x)|^2 + |g_2(x)|^2)^{1/2}\|_\infty$. Vese and Osher suggested a variational minimization problem that approximate the above model. This approach essentially searches for the solution u, g_1, g_2 of

$$\inf_{(u,\, g_1,\, g_2)} \{\|u\|_{BV} + \lambda \|v\|_{BV^*}\} \quad \text{subject to } f = u + v. \quad (12)$$

A numerical algorithm to solve this problem is described in [3] with encouraging simulation results. Since the direct treatment of the BV^* in the above formulation is hard, Vese and Osher proposed an approximation by using an L^p norm of the g_1, g_2 functions. Also, the constraint is replaced by a penalty of the form $\mu\|f - u - v\|_2^2$. Their method approaches Meyer's model as p and μ go to infinity.

Although the approach we take is totally different, it bares some similarities in spirit to the above described method. Referring to our formulation in (11) with the choice $\gamma = 0$

$$\min_{\{\underline{X}_t,\, \underline{X}_n\}} \left\|\mathbf{T}_n^+ \underline{X}_n\right\|_1 + \left\|\mathbf{T}_t^+ \underline{X}_t\right\|_1 + \lambda \|\underline{X} - \underline{X}_t - \underline{X}_n\|_2^2 \quad (13)$$

we see the following connections (note that equivalence is not claimed here).

- Based on our previous discussion on the relation between the TV and the undecimated Haar, we can propose $\|\mathcal{H}u\|_1$ as a replacement to $\|u\|_{BV}$. Here, $\mathcal{H}$ is the undecimated Haar transform (i.e., $\mathcal{H} = T_n^+$ in our original notations). Thus, there is a similarity between the effects of the first terms in both (12) and (13).
- We may argue that images with sparse representations in the DCT domain (local with varying block sizes and block overlap) present strong oscillations and, therefore, could be considered as textures, belonging to the Banach space $BV^*(\mathcal{R}^2)$. This suggests that $\|v\|_{BV^*}$ could also be approximated by an ℓ^1 norm term $\|\mathcal{D}v\|_1$ where $\mathcal{D}$ is the DCT transform (i.e., $\mathcal{D} = T_t^+$ in our notations). This leads to a similarity between the second terms in the two optimization problems (12) and (13).
- The third expression is exactly the same in (12) and (13), after the Vese–Osher modifications. Thus, we see a close relation between our model and the one proposed by Meyer as adopted and used by Vese and Osher. However, there are also major differences that should be mentioned.
- In our implementation, we do not use the undecimated Haar with just one resolution, but rather use the complete pyramid. We should note that The variational approach could be extended to have a multiscale treatment by adopting spatially adaptive and resolution adaptive coefficient λ.
- We have replaced the Haar with more effective transforms such as curvelet. Several reasons justify such a change. Among them is the fact that curvelet better succeeds in detecting noisy edges.
- Our method does not search for the implicit g_1, g_2 supposed to be the origin of the texture, but rather searches directly the texture part by an alternative and simpler model based on the local DCT.
- We should note that the methodology presented in the paper is not limited to the separation of texture and piecewise-smooth parts of an image. The basic idea advocated here is how to separate signals to different content types, leaning on the idea that each of the ingredients have a sparse representation with a proper choice of a dictionary. This may lead to other applications and different implementations. We leave this generalized view for future research.
- As a final note, we should remark that the Vese–Osher technique is much faster than the one presented here. The prime reason for this gap is the curvelet transform runtime. Future versions of curvelet may change this shortcoming.

B. Compression via Separation

A pioneering work described in [2] proposes a separation of cartoon from texture for efficient image compression. This algorithm relies on an experience gained on similar decompositions applied to audio signals [37]. Our algorithm is very similar in spirit to the approach taken in [2], namely, use of different dictionaries for effective (sparse) representation of each content type, and pursuit that seeks the sparsest of all representations. Still there are several major differences worth mentioning.

- While our algorithm uses curvelet, ridgelet, and several other types of over-complete transforms, the chosen dictionaries in [2] are confined to be orthonormal wavelet packets (optimized per the task). This choice is crucial for the compression to follow, but causes loss of sparsity in the representations.
- Our separation approach is parallel, seeking jointly a decomposition of the image into the two ingredients. The numerical implementation uses "Sardy-like" sequential transforms followed by soft thresholding, but applied iteratively, the algorithm gets closer to the BP result, which is essentially a parallel decomposition technique. The algorithm in [2] is sequential, pealing the cartoon content and then treating the reminder as texture.
- The proposed method in [2] concentrates on compression performance, and has less interest in the visual quality of the separation. The algorithm presented here, on the other hand, is all about getting pleasing images to a human viewer. This is why TV penalty was added to treat ringing artifacts.
- A large portion of our work came as a direct consequence to the theoretical study we have done on the BP performance limits (see Appendix II). When we assume sparsity under the chosen dictionaries, we can invoke the uniqueness result, that says that the original sparsity pattern is indeed the sparsest one possible. When we employ the BP for numerically getting the result, we lean on the

equivalence result promising that if indeed the combination is sparse enough, BP will find it well. The work in [2] claims of success are leaning on the actual obtained compression results.

Very recent similar attempt to exploit separation for image compression is reported in [5]. The authors use the variational paradigm for achieving the separation, and then consider compression of each content type separately, as in [2].

The separation algorithm presented in [4] is proposed for a general analysis of image content and not compression. However, it bares some similarities to both the algorithm in [2] and the one presented in this paper. As in [2], the decomposition of the image content is sequential: The first stage extracts the sketchable content (similar to the piecewise smooth content, but different), and this is achieved by the matching pursuit algorithm, applied with a trained dictionary of local primitives. The second stage represents the nonsketchable (texture) content and is based on Markov random field (MRF) representation. The goal of the proposed separation in [4] is somewhat different than the one discussed here, as it focuses on a sparse description of the sketched image. This is in contrast to the method proposed here where sparsity is desired and found across all content types.

VII. Discussion

In this paper, we have presented a novel method for separating an image into its texture and piecewise smooth ingredients. Our method is based on the ability to represent these content types as sparse combinations of atoms of predetermined dictionaries. The proposed approach is a fusion of the BP algorithm and the TV regularization scheme, both merged in order to direct the solution toward a successful separation.

This paper offers a theoretical analysis of the separation idea with the BP algorithm, and shows that a perfect decomposition of image content could be found in principle. While the theoretical bounds obtained for a perfect decomposition are rather weak, they serve both as a starting point for future research, and as motivating results for the practical sides of the work.

In going from the pure theoretic view to the implementation, we manage to extend the model to treat additive noise—essentially any content in the image that does not fit well with either texture or piecewise-smooth contents. We also chang the problem formulation, departing from the BP, and getting closer to a maximum *a posteriori* estimation method. The new formulation leads to smaller memory requirements, and the ability to add helpful constraints.

Appendix I
Block-Coordinate-Relaxation Method

In Section II-C, we have seen an alternative formulation to the separation task, built on the assumption that the involved dictionaries are concatenations of unitary matrices. Thus, we need to minimize (7), given (after a simplification) as

$$\min_{\{\underline{\alpha}(k)\}_{k=1}^{L}} \quad \sum_{k=1}^{L} \|\underline{\alpha}(k)\|_1 + \lambda \left\| \underline{X} - \sum_{k=1}^{L} \mathbf{T}(k)\underline{\alpha}(k) \right\|_2^2. \tag{AI1}$$

Note that we have discarded the TV part for the discussion given here. We also simply assume that the unknowns $\underline{\alpha}(k)$ contain both the texture and the piecewise-smooth parts.

Minimizing such a penalty function was shown by Bruce, Sardy, and Tseng [22] to be quite simple, as it is based on the shrinkage algorithm due to Donoho and Johnston [21]. In what follows, we briefly describe this algorithm and its properties.

1) Property 1: Referring to (AI1) as a function of $\{\underline{\alpha}(k)\}_{k_0}$, assuming all other unknowns as known, there is a closed-form solution for the optimal $\{\underline{\alpha}(k)\}_{k_0}$, given by

$$\{\underline{\alpha}(k)\}_{k_0}^{\text{opt}} = \text{sign}(\underline{Z}) \cdot \left(|\underline{Z}| - \frac{1}{2\lambda} \right)_+ \tag{AI2}$$

for $\underline{Z} = \underline{X} - \sum_{k=1,\ k\neq K_0}^{L} \mathbf{T}(k)\underline{\alpha}(k)$.

This property is the source of the simple numerical scheme of the block-coordinate-relaxation method. The idea is to sweep through the vectors $\underline{\alpha}(k)$ one at a time repeatedly, fixing all others, and solving for each.

2) Property 2: Sweeping sequentially through k and updating $\underline{\alpha}(k)$ as in Property 1, the block-coordinate-relaxation method is guaranteed to converge to the optimal solution of (AI1).

Appendix II
Theoretic Analysis of the Separation Task

In this Appendix, we aim to show that the separation as described in this paper has strong theoretical justification roots. Those lean on some very recent results in the study of the BP performance. The presented material in this appendix is deliberately brief, with the intention to present a more extensive theoretical study in a separate paper.

We start with (3) that stands as the basis for the separation process. This equation could also be written differently as

$$\underline{\alpha}_{\text{all}}^{\text{opt}} = \begin{bmatrix} \underline{\alpha}_t^{\text{opt}} \\ \underline{\alpha}_n^{\text{opt}} \end{bmatrix} = \text{Arg} \min_{\{\underline{\alpha}_t,\ \underline{\alpha}_n\}} \left\| \begin{bmatrix} \underline{\alpha}_t \\ \underline{\alpha}_n \end{bmatrix} \right\|_0$$
$$\text{subject to : } \underline{X} = [\,\mathbf{T}_t \quad \mathbf{T}_n\,] \begin{bmatrix} \underline{\alpha}_t \\ \underline{\alpha}_n \end{bmatrix} = \mathbf{T}_{\text{all}}\underline{\alpha}_{\text{all}}. \tag{AII1}$$

From [14], we recall the definition of the *Spark*.

Definition 1: Given a matrix $\mathbf{A}$, its Spark ($\sigma_A = Spark\{\mathbf{A}\}$) is defined as the minimal number of columns from the matrix that form a linearly dependent set.

Based on this we have the following result in [14] that gives a guarantee for global optimum of (AII1) based on a sparsity condition.

Theorem 1: If a candidate representation $\underline{\alpha}_{\text{all}}$ satisfies $\|\underline{\alpha}_{\text{all}}\|_0 < Spark\{\mathbf{T}_{\text{all}}\}/2$, then this solution is necessarily the global minimum of (AII1).

Based on this result it is clear that the higher the value of the *Spark*, the stronger this result is. Immediate implication from the above is the following observation, referring to the success of the separation process.

Corollary 1: If the image $\underline{X} = \underline{X}_t + \underline{X}_n$ is built such that $\underline{X}_t = \mathbf{T}_t\underline{\alpha}_t$ and $\underline{X}_n = \mathbf{T}_n\underline{\alpha}_n$, and $\|\underline{\alpha}_t\|_0 + \|\underline{\alpha}_n\|_0 < Spark\{\mathbf{T}_{\text{all}}\}/2$ is true, then the global minimum of (AII1) is necessarily the desired separation.

Proof: The proof is simple deduction from Theorem 1.

Actually, a stronger claim could be given if we assume a successful choice of dictionaries $\mathbf{T}_t$ and $\mathbf{T}_n$. Let us define a variation of the *Spark* that refers to the interface between atoms from two dictionaries.

Definition 2: Given two matrices $\mathbf{A}$ *and* $\mathbf{B}$ *with the same number of rows, their Inter-Spark* ($\sigma_{A\leftrightarrow B} = Spark\{\mathbf{A},\mathbf{B}\}$) *is defined as the minimal number of columns from the concatenated matrix* $[\mathbf{A},\mathbf{B}]$ *that form a linearly dependent set, and such that columns from both matrices participate in this combination.*

An important feature of our problem is that the goal is the successful separation of content of an incoming image and not finding the true sparse representation per each part. Thus, a stronger claim can be made.

Corollary 2: Suppose the image $\underline{X} = \underline{X}_t + \underline{X}_n$ *is built such that* $\underline{X}_t = \mathbf{T}_t\underline{\alpha}_t$ *and* $\underline{X}_n = \mathbf{T}_n\underline{\alpha}_n$. *If* $\|\underline{\alpha}_t\|_0 + \|\underline{\alpha}_n\|_0 < \sigma_{\mathbf{T}_t\leftrightarrow\mathbf{T}_n}/2$ *and* $\|\underline{\alpha}_t\|_0, \|\underline{\alpha}_n\|_0 > 0$ *(i.e., there is a mixture of the two), then if the global minimum of (AII1) satisfies* $\|\underline{\alpha}_t^{\text{opt}}\|_0, \|\underline{\alpha}_n^{\text{opt}}\|_0 > 0$, *it is necessarily the successful separation.*

Proof: Given a mixture of columns from the two dictionaries, by the definition of the *Inter-Spark* it is clear that if there are fewer than $\sigma_{\mathbf{T}_t\leftrightarrow\mathbf{T}_n}/2$ nonzeros in such combination, it must be the unique sparsest solution. The new bound is higher than $Spark\{\mathbf{T}_{\text{all}}\}/2$, and, therefore, this result is stronger.

So far, we concentrated on (AII1) which stands as the ideal (but impossible) tool for the separation. An interesting question is why should the ℓ^1 replacement succeed in the separation as well. In order to answer this question we have to define first the *Mutual Incoherence*.

Definition 3: Given a matrix $\mathbf{A}$, *its* $Mutual - Incoherence\{\mathbf{A}\} = M_A$ *is defined as the maximal off-diagonal entry in the absolute Gram matrix* $|\mathbf{A}^H\mathbf{A}|$.

The *mutual incoherence* is closely related to the *Spark*, and, thus, one can similarly define a similar notion of *inter–*M_A. We have the following result in [14].

Theorem 2: If the solution $\underline{\alpha}_{\text{all}}^{\text{opt}}$ *of (AII1) satisfies* $\|\underline{\alpha}_{\text{all}}^{\text{opt}}\|_0 < (1/M_{\mathbf{T}_{\text{all}}} + 1)/2$, *then the* ℓ^1 *minimization alternative is guaranteed to find it.*

For the separation task, this Theorem implies that the separation via (4) is successful if it is based on sparse enough ingredients.

Corollary 3: If the image $\underline{X} = \underline{X}_t + \underline{X}_n$ *is built such that* $\underline{X}_t = \mathbf{T}_t\underline{\alpha}_t$ *and* $\underline{X}_n = \mathbf{T}_n\underline{\alpha}_n$, *and* $\|\underline{\alpha}_t\|_0 + \|\underline{\alpha}_n\|_0 < (1/M_{\mathbf{T}_{\text{all}}} + 1)/2$ *is true, then the solution of (4) leads to the global minimum of (AII1) and this is necessarily the desired separation.*

Proof: The proof is simple deduction from Theorem 2.

We should note that the bounds given here are quite restrictive and does not reflect truly the much better empirical results. The above analysis is coming form a *worst-case* point of view (e.g., see the definition of the *Spark*), as opposed to the average case we expect to encounter empirically. Nevertheless, the ability to prove perfect separation in a stylized application without noise and with restricted success is of great benefit as a proof of concept.

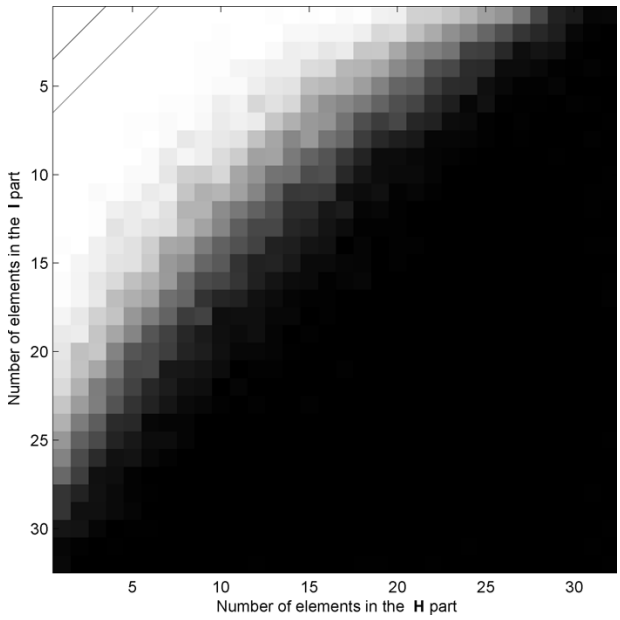

Fig. 9. Empirical probability of success of the BP algorithm for separation of sources. Per every sparsity combination, 100 experiments are performed and the success rate is computed. Theoretical bounds are also drawn for comparison.

In order to demonstrate the gap between theoretical results and empirical evidence in BP separation performance, Fig. 9 presents a simulation of the separation task for the case of signal $\underline{X}$ of length 64, a dictionary built as the combination of the Hadamard unitary matrix (assumed to be $\mathbf{T}_t$) and the identity matrix (assumed to be $\mathbf{T}_n$). We randomly generate sparse representations with varying number of nonzeros in the two parts of the representation vector (of length 128), and present the empirical probability (based on averaging 100 experiments) to recover correctly the separation.

For this case, Corollary 3 suggests that the number of nonzero in the two parts should be smaller than $0.5 \cdot (1 + 1/M) = (1 + \sqrt{64})/2 = 4.5$. Actually a better result exists for this case in [15] due to the construction of the overall dictionary as a combination of two unitary matrices. Thus, the better bound is $(\sqrt{2} - 0.5))/M = 7.3$. Both these bounds are overlayed on the empirical results in the figure, and as can be seen, BP succeeds well beyond the bound. Moreover, this trend is expected to strengthen as the signal size grows, since than the worst-case scenarios (for which the bounds refer to) become of smaller probability and of less affect on the average result.

It is interesting to note that very recent attempts by several research groups managed to quantify the average behavior of the BP in probabilistic terms. A pioneering work by Candes and Romberg [38] established one such important result, and several others follow, although none are published yet.

ACKNOWLEDGMENT

The authors would like to thank Prof. S. Osher and Prof. L. Vese for helpful discussions and for sharing their results to be presented in this paper.

REFERENCES

[1] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neur. Comput.*, vol. 13, pp. 863–882, 2001.

[2] F. Meyer, A. Averbuch, and R. Coifman, "Multilayered image representation: Application to image compression," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1072–1080, Sep. 2002.

[3] L. Vese and S. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. Sci. Comput.*, vol. 19, pp. 553–577, 2003.

[4] C. Guo, S. Zhu, and Y. Wu, "A mathematical theory of primal sketch and sketchability," presented at the 9th IEEE Int. Conf. Computer Vision, Nice, France, Oct. 2003.

[5] J. Aujol and B. Matei, "Structure and texture compression," INRIA Project ARIANA, Sophia Antipolis, France, Tech. Rep. ISRN I3S/RR-2004-02-FR, 2004.

[6] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 882–889, Aug. 2003.

[7] J. Aujol, G. Aubert, L. Blanc-Feraud, and A. Chambolle, "Image decomposition: Application to textured images and SAR images," INRIA Project ARIANA, Sophia Antipolis, France, Tech. Rep. ISRN I3S/RR-2003-01-FR, 2003.

[8] J. Aujol and A. Chambolle, "Dual norms and image decomposition models," INRIA Project ARIANA, Sophia Antipolis, France, Tech. Rep. ISRN 5130, 2004.

[9] Y. Meyer, "Oscillating patterns in image processing and non linear evolution equations," in *Univ. Lecture Ser.*, 2002, vol. 22, AMS.

[10] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation noise removal algorithm," *Phys. D*, vol. 60, pp. 259–268, 1992.

[11] S. Chen, D. Donoho, and M. Saunder, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.

[12] J.-L. Starck, E. Candes, and D. Donoho, "Astronomical image representation by the curvelet tansform," *Astron. Astrophys.*, vol. 398, pp. 785–800, 2003.

[13] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

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

[15] M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[16] R. Gribonval and M. Nielsen, "Some remarks on nonlinear approximation with schauder bases," *East J. Approx.*, vol. 7, no. 2, pp. 267–285, 2001.

[17] J.-L. Starck, D. Donoho, and E. Candès, "Very high quality image restoration," presented at the 9th SPIE Conf. Signal and Image Processing: Wavelet Applications in Signal and Image Processing, A. Laine, M. Unser, and A. Aldroubi, Eds., San Diego, CA, Aug. 2001.

[18] E. Candès and F. Guo, "New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction," *Signal Process.*, vol. 82, no. 5, pp. 1516–1543, 2002.

[19] J.-L. Starck, M. Nguyen, and F. Murtagh, "Wavelets and curvelets for image deconvolution: A combined approach," *Signal Process.*, vol. 83, no. 10, pp. 2279–2283, 2003.

[20] F. Malgouyres, "Minimizing the total variation under a general convex constraint for image restoration," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1450–1456, Dec. 2002.

[21] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[22] A. Bruce, S. Sardy, and P. Tseng, "Block coordinate relaxation methods for nonparametric signal de-noising," *Proc. SPIE*, vol. 3391, pp. 75–86, 1998.

[23] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[24] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[25] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.

[26] E. Candès and D. Donoho, "Ridgelets: The key to high dimensional intermittency?," *Phil. Trans. Roy. Soc. London A*, vol. 357, pp. 2495–2509, 1999.

[27] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[28] J.-L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*. New York: Springer-Verlag, 2002.

[29] E. J. Candès, "Harmonic analysis of neural netwoks," *Appl. Comput. Harmon. Anal.*, vol. 6, pp. 197–218, 1999.

[30] J.-L. Starck, E. Candès, and D. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 131–141, Jun. 2002.

[31] D. Donoho and M. Duncan, "Digital curvelet transform: Strategy, implementation and experiments," in *Proc. Aerosense Wavelet Applications VII*, vol. 4056, H. Szu, M. Vetterli, W. Campbell, and J. Buss, Eds., 2000, pp. 12–29.

[32] E. J. Candès and D. L. Donoho, "Curvelets—A surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting: Saint-Malo 1999*, A. Cohen, C. Rabut, and L. Schumaker, Eds. Nashville, TN: Vanderbilt Univ. Press, 1999.

[33] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides," Dept. Math., Univ. Bremen, Bremen, Germany, Tech. Rep. 26, 2003.

[34] E. J. Candès and D. L. Donoho, "Recovering ddges in ill-posed inverse problems: Optimality of curvelet frames," Tech. Rep., Dept. Stat., Stanford Univ., Stanford, CA, 2000.

[35] M. Vetterli, "Wavelets, approximation, and compression," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 59–73, Sep. 2001.

[36] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Texture preserving variational denoising using an adaptive fidelity term," in *Proc. VLSM*, Nice, France, 2003, pp. 137–144.

[37] R. Coifman and F. Majid, "Adapted waveform analysis and denoising," in *Progress in Wavelet Analysis and Applications*, Frontières ed., Y. Meyer and S. Roques, Eds., 1993, pp. 63–76.

[38] E. Candès and J. Romberg, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," private communication, 2004.

Jean-Luc Starck received the Ph.D. degree from the University Nice-Sophia Antipolis, France, and the Habilitation degree from the University Paris XI, Paris, France.

He was a Visitor at the European Southern Observatory (ESO) in 1993 and at the Statistics Department, Stanford University, Stanford, CA, in 2000. He has been a Researcher at CEA, Gif sur Yvette, France, since 1994. He is the author of two books entitled *Image Processing and Data Analysis: the Multiscale Approach* (Cambridge, MA: Cambridge Univ. Press, 1998) and *Astronomical Image and Data Analysis* (New York: Springer, 2002). His research interests include image processing, multiscale methods, and statistical methods in astrophysics.

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

From 1988 to 1993, he served in the Israeli Air Force. From 1997 to 2000, he worked at Hewlett-Packard Laboratories, Israel, as an R&D Engineer. From 2000 to 2001, he headed the research division at Jigami Corporation, Israel. From 2001 to 2003, he was a Research Associate with the Computer Science Department, Stanford University (SCCM program), Stanford, CA. In September 2003, he joined the Department of Computer Science, IIT, as an Assistant Professor. He was also a Research Associate at IIT from 1998 to 2000, teaching courses in the Electrical Engineering Department. He works in the field of signal and image processing, specializing, in particular, on inverse problems, sparse representations, and over-complete transforms.

Dr. Elad received the Best Lecturer Award twice (in 1999 and 2000). He is also the recipient of the Guttwirth and the Wolf fellowships.

David L. Donoho received the B.A. degree (summa cum laude) in statistics from Princeton University, Princeton, NJ, where his senior thesis adviser was J. W. Tukey, and the Ph.D. degree in statistics from Harvard University, Cambridge, MA, where his Ph.D. adviser was P. Huber.

He is a Professor of statistics, Stanford University, Stanford, CA. He was previously a Professor at the University of California, Berkeley, and a Visiting Professor at the Universite de Paris, Paris, France, as well as a Sackler Fellow at Tel Aviv University, Tel Aviv, Israel. His research interests are in harmonic analysis, image representation, and mathematical statistics.

Dr. Donoho is a member of the USA National Academy of Sciences and a Fellow of the American Academy of Arts and Sciences.