A Review of Adaptive Image Representations — **Source link** ⧉

Gabriel Peyré

Institutions: Paris Dauphine University

Related papers:

- $rm K$ -SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation

- Nonlinear total variation based noise removal algorithms

- Image quality assessment: from error visibility to structural similarity

- Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries

- Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering

A Review of Adaptive Image Representations

Gabriel Peyré

A Review of Adaptive Image Representations

Gabriel Peyré

Abstract—**Improving the modeling of natural images is important to go beyond the state of the art for many image processing tasks such as compression, denoising, inverse problems and texture synthesis. Natural images are composed of intricate patterns such as regular areas, edges, junctions, oriented oscillations and textures. Processing efficiently such a wide range of regularities requires methods that are adaptive to the geometry of the image. This adaptivity can be achieved using sparse representations in a redundant dictionary. The geometric adaptivity is important to search for efficient representations in a structured dictionary. Another way to capture this geometry is through non-local interactions between patches in the image. The resulting non-local energies can be used to perform an adaptive image restoration. This paper reviews these emerging technics and shows the interplay between sparse and non-local regularizations.**

Index Terms—**Adaptivity, sparse regularization, non-local regularization, best basis, triangulations, dictionary learning, approximation, denoising, inverse problems, texture synthesis.**

Finding efficient representations for natural images is relevant for many image processing tasks. Sparse and non-local regularizations have recently emerged as unifying concepts to formalize this notion of geometric representations. These methods have been proved useful to perform compression, denoising, inversion of linear operators and texture synthesis. This paper reviews these technics and shows that the concepts of sparsity and variational minimization are at the heart of the success of these methods.

I. INTRODUCTION

A large class of image processing problems can be formalized as recovering a high resolution image $f^\star \in \mathbb{R}^N$ of N pixels from noisy observations

$y \in \mathbb{R}^Q$, where $Q \leqslant N$. The forward model assumes that y is a degraded version of an unknown high resolution image f^0, where $y = \Phi f^0 + w$. The linear operator $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^Q$ models the acquisition process.

If $\Phi = \mathrm{Id}$, the recovery of a good approximation of f^0 corresponds to removing noise. When Φ is ill-posed, the inverse problem of recovering a high resolution image is more difficult. Typical inverse problems are super-resolution, inpainting (see for instance [64], [5])and compressed sensing [11], [33].

Non-adaptive regularization. Classical regularization theory makes use of a prior $J(f) \in \mathbb{R}$ that is intended to be low for the images one is interested in. Priors based on derivatives of the image impose some smoothness on the recovered image. The simplest prior is the ℓ^2 norm of the gradient $J(f) = \sum_x \|\nabla f(x)\|^2$, where ∇f is a finite difference approximation of the gradient. This corresponds to a Sobolev prior that favors uniformly smooth images. The total variation prior, introduced for denoising by Rudin, Osher and Fatemi [84], reads $J(f) = \sum_x \|\nabla f(x)\|$. It allows one to take into account discontinuous images, and favors objects in the image with a small perimeter. Other classes of priors include sparsity in orthogonal or redundant representations, which are discussed in Section II.

A prior is taken into account during the inversion of $y = \Phi f^0 + w$ using a variational minimization

$$f^\star \in \operatorname*{argmin}_{f \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi f\|^2 + \mu J(f), \quad (1)$$

where μ weights the fidelity to the observations y versus the regularity of the solution, and should be adapted to the noise level.

Adaptive regularization. In this paper, the adaptation to the structures of the image is performed through the optimization of some geometric parameter λ. This parameter takes different forms, such as a quadtree in Section III, an association field in Section IV-B, a triangulation in Section

IV-C or a weighted graph in Section V. This parameter always caries some geometrical information about the location of edges or the direction of textures.

This geometry λ parameterizes a family of prior $J^\lambda(f)$ that favors images that are smooth according to the estimated geometry. The non-adaptive regularization (1) is turned into an adaptive recovery by minimizing

$$(f^\star, \lambda^\star) \in \underset{f \in \mathbb{R}^N, \lambda}{\operatorname{argmin}} E(f, \lambda) \quad \text{where} \quad (2)$$

$$E(f, \lambda) = \frac{1}{2}\|y - \Phi f\|^2 + \mu J^\lambda(f) + \mathcal{E}(\lambda).$$

The additional term $\mathcal{E}(\lambda)$ allows one to introduce constraints on the set of admissible geometries.

For a fixed λ, the energy $E(f, \lambda)$ is usually convex, but E is non-convex jointly on (f, λ). One thus has to resort to iterative schemes that progressively estimate both the image and the geometry, and converge to a stationary point $(f^\star, \lambda^\star)$ of E.

The following sections describe several instances of this adaptive regularization framework (2), each time using a different parameterization of the prior J^λ adapted to several kinds of edge or texture features.

II. SPARSE AND ADAPTIVE REPRESENTATIONS

Computing a sparse decomposition corresponds to approximating an image $f^0 \in \mathbb{R}^N$ of N pixels as $f_M = \sum_{m=0}^{P-1} a_m^\star \psi_m$ using a small number $M = \|a^\star\|_0 = |\{m \setminus a_m^\star \neq 0\}|$ of atoms chosen in a dictionary $\mathcal{D} = \{\psi_m\}_m \subset \mathbb{R}^N$ of $P \geqslant N$ atoms.

A. Non-linear Approximation

Best approximation and thresholding. A sparse approximation is efficient if the error $\|f^0 - f_M\|$ is small. The best approximation f_M with M non-zero coefficient in the dictionary $\mathcal{D} = \{\psi_m\}_m$ can be found by solving a variational minimization $f_M = \sum_{m=0}^{P-1} a_m^\star \psi_m$ where

$$a^\star \in \underset{a \in \mathbb{R}^P}{\operatorname{argmin}} \|f^0 - \sum_{m=0}^{P-1} a_m \psi_m\|^2 + T^2\|a\|_0 \quad (3)$$

In the case where $\mathcal{D} = \mathcal{B} = \{\psi_m\}_m$ is an orthogonal basis of $P = N$ atoms, the problem (3) is re-written as

$$f_M \in \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \|f^0 - f\|^2 + T^2 J(f) \quad (4)$$

where $\quad J(f) = |\{m \setminus \langle f, \psi_m \rangle \neq 0\}|$

for a value of T adapted so that $J(f) = M$. This is the variational regularization (1) for the identity operator $\Phi = \text{Id}$, no noise $w = 0$, $2\mu = T^2$, and the ℓ^0 prior J. Although this prior J is non-convex, a global minimizer f_M is found using a hard thresholding that selects the inner products with largest amplitudes, see for instance [61]

$$f_M = H_T(f^0, \mathcal{B}) = \sum_{|\langle f^0, \psi_m \rangle| > T} \langle f^0, \psi_m \rangle \psi_m,$$

$$(5)$$

$$\text{and} \quad M = |\{m \setminus |\langle f^0, \psi_m \rangle| > T\}|$$

Computing sparse approximations f_M in orthogonal bases is at the heart of many efficient compression algorithms. For instance JPEG-2000 uses a nearly orthogonal wavelet basis and is a state of the art coding scheme, see [61].

Decay of approximation error. For several classes of images, one can prove the asymptotic optimality of such a non-linear approximation scheme (5) in well chosen ortho-bases. Uniformly C^α images enjoy an optimal asymptotic approximation error decay $\|f^0 - f_M\|^2 = O(M^{-\alpha})$ in a Fourier basis. Discontinuous images with bounded variations, that have edges of finite total perimeter, enjoy an optimal asymptotic decay $\|f^0 - f_M\|^2 = O(M^{-1})$ in a wavelet basis. Figure 1 shows some examples of smooth and discontinuous images.

Wavelets are however not optimal for piecewise regular images whose edges are regular. The error for such a cartoon image decays as $O(M^{-1})$. The square support of wavelet atoms forbid them to capture the directionality of regular edges. The curvelet frame of Candès and Donoho [10] enjoys an error decay of $O(\log(M)^3 M^{-2})$ for images that are C^2 outside C^2 edges, thus enhancing over the wavelet approximation. A discrete implementation of the curvelet transform corresponds to the projection of the image $f \in \mathbb{R}^N$ on a tight frame [12]. This curvelets approximation has also been used for inverse problems regularization [13]. For smoothly varying textures, a fixed waveatom frame [28]provides a small approximation error.

To approximate efficiently images with complicated edges and textures, more redundant dictionaries are needed. To keep the computation of the approximation tractable, it is however important to use a structured dictionary. Structuring the dictionary is also important for compression in order to reduce the coding cost.
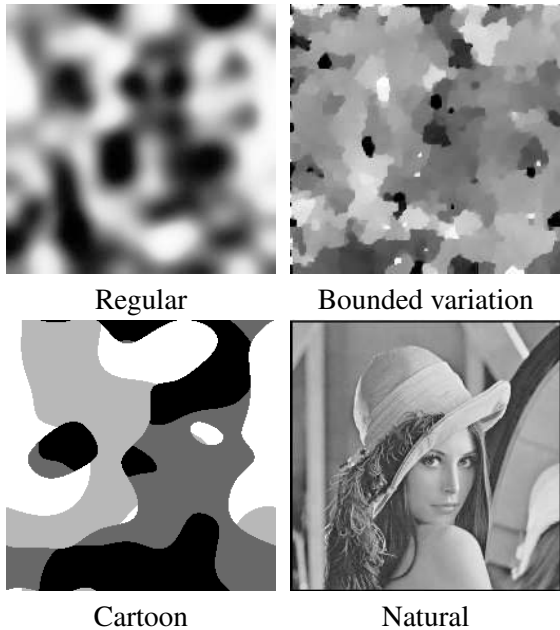
Fig. 1. *Examples of images within several complexity classes.*

B. Sparse Approximation over Redundant Dictionaries

For an arbitrary redundant dictionary, computing the best M-term approximation (3) is however NP-hard [27], and one thus needs to resort to greedy schemes [63] or ℓ^1 minimization.

ℓ^1 **sparse approximation.** A sparse approximation is obtained by convexifying the ℓ^0 pseudo norm in (3), and solving the following basis pursuit denoising convex problem [15], $f^\star = \sum_{m=0}^{P-1} a_m^\star \psi_m$ where

$$a^\star \in \operatorname*{argmin}_{a \in \mathbb{R}^P} \frac{1}{2}\|f^0 - \sum_{m=0}^{P-1} a_m \psi_m\| + \mu\|a\|_1, \quad (6)$$

where $\mu > 0$ is adapted so that $\|a^\star\|_0 = M$ and where the ℓ^1 norm is defined as $\|a\|_1 = \sum_{m=0}^{P-1} |a_m|$. Many algorithms have proposed to solve the convex problem (6), for instance iterative thresholding methods [25], [24]. Sparse coding in redundant dictionaries finds application to compression of sounds [51] and images [45].

Such ℓ^1 sparse approximation can be shown to be close to optimal for a restricted class of dictionaries, see for instance [92]. Most image processing dictionaries are however too redundant to fit within this theoritical analysis. Furthermore, for highly redundant dictionaries, solving (6) is computationally demanding.

Redundant and structured dictionaries. To avoid these pitfalls, one needs to use structured dictionaries which enable a theoritical analysis of the approximation performance for restricted class of images, and also enable a fast search for nearly optimal approximations. Sections III and IV detail several structured dictionaries, together with adaptive strategies to perform this search.

C. Sparse Approximation for Processing

Building ortho-bases or frames for which the approximation error decays fast is important for many applications. One can show that for several image processing problems, the efficiency of the processing is related to the non-linear approximation error. This section only sketches the main results for compression, denoising, and inverse problem regularization. It gives informal statements that require additional technical hypotheses to be mathematically rigorous.

Compression. Efficient image compression schemes perform a binary coding of quantized coefficients $\langle f^0, \psi_m \rangle$ where $\{\psi_m\}_m$ is an orthogonal basis. An image $f^0 \in \mathbb{R}^N$ is coded with R bits, and the decoder computes an approximation f_R of f^0.
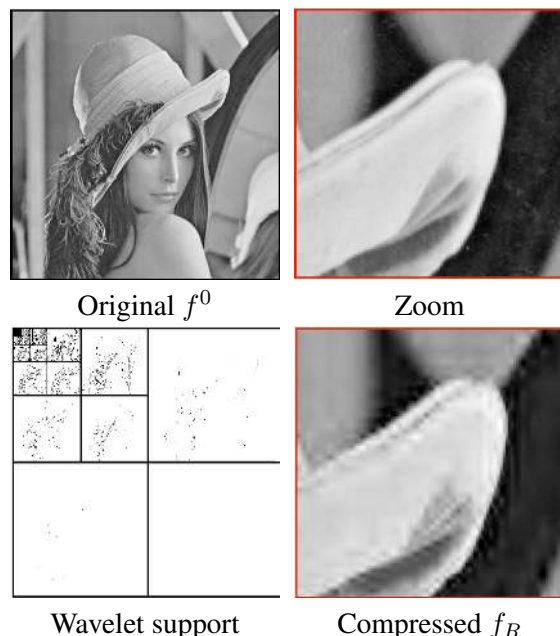

Fig. 2. *Image compression using wavelet support codding.*

To exploit the sparsity of the decomposition of the image, a support coding scheme codes in-

dependently the locations of the non-zero coefficients, and their amplitudes. Figure 2 shows an example of compression using this simple coding scheme.

Under some conditions on the basis, which are satisfied by many classical ortho-bases such as wavelets, one can shows that if the sorted magnitudes $|\langle f^0, \psi_m \rangle|$ decay like $m^{-\frac{\alpha-1}{2}}$, then the coding-decoding error satisfies

$$\|f^0 - f_R\|^2 = O(R^{-\alpha} \log^{\alpha}(R)), \qquad (7)$$

see [61]. Note that this decay of the coefficients implies $\|f^0 - f_M\|^2 = O(M^{-\alpha})$. This shows that a basis that is efficient to perform non-linear image approximation is able to efficiently perform image compression by support coding. State of the art coders such as JPEG-2000 [61] use advanced conditional coding schemes that exploit statistical redundancies among the coefficients. This improves the decay of the distortion $\|f^0 - f_R\|$ with respect to a naive support coder.

This result does not extend to an arbitrary dictionary, since it might be too costly to code the coefficients. It extends to dictionaries of ortho-bases detailed in Section (III), if the number of atoms within the dictionary is not too large.

Denoising. Denoising corresponds to estimating an image from a noisy observation $y = f^0 + w$, where $f^0 \in \mathbb{R}^N$ is a deterministic, unknown, clean image. A standard assumption is that w is a Gaussian white noise of variance σ^2.

A popular non-linear denoising method is the hard thresholding $f^* = H_T(y, \mathcal{B})$ over an orthogonal basis $\mathcal{B}$, where H_T is defined in (5). This corresponds to the minimization of (1) for $\Phi = \mathrm{Id}$ and J the ℓ^0 prior. It was introduced by Donoho and Johnstone [35], and uses a well chosen T proportional to the noise level σ. In an orthogonal basis $\mathcal{B}$, the decay $\|f^0 - f_M\|^2 = O(M^{-\alpha})$ implies

$$E_w(\|f^0 - H_T(y, \mathcal{B})\|^2) = O(\sigma^{\frac{2\alpha}{\alpha+1}}), \quad (8)$$

if one uses the universal threshold $T = \sigma\sqrt{2\log(N)}$, see [35]. Here E_w denotes the expectation with respect to the noise random variable w, and f_M is the best M terms approximation of f^0 in $\mathcal{B}$. This shows that the average denoising error decays fast to zero with σ if $\mathcal{B}$ is able to efficiently approximate f^0.

Denoising by thresholding is extended to redundant frames, where f^* is reconstructed using the dual frame. Denoising artifacts are reduced by

using a translation invariant redundant frame [23]. Denoising of images is also enhanced by using statistical models that take into account the non-independence of coefficients of natural images, see for instance [83].

Inverse problems. For inverse problems, one measures $y = \Phi f^0 + w$, where w is an additive noise, and where $\Phi : \mathbb{R}^N \to \mathbb{R}^Q$ is ill-conditionned. To recover an approximation f^* of f^0 from y, one replaces the sparse approximation by a ℓ^1 regularization $f^* = \sum_{m=0}^{P-1} a_m^* \psi_m$ where

$$a^* \in \operatorname*{argmin}_{a \in \mathbb{R}^P} \frac{1}{2} \|y - \Phi \sum_{m=0}^{P-1} a_m \psi_m\|^2 + \mu \sum_{m=0}^{P-1} |a_m|. \tag{9}$$

This is equivalent to computing a sparse decomposition of y in the redundant dictionary $\Phi\mathcal{D} = \{\Phi\psi_m\}_{m=0}^{P-1}$ of $\mathbb{R}^Q$ using basis pursuit denoising (6).

In the special case where $\mathcal{D} = \mathcal{B}$ is an orthogonal basis, the minimization (9) can be re-written as the variational problem (1) using an ℓ^1 prior $J(f) = \sum_m |\langle f, \psi_m \rangle|$.

A few theoritical results exist, but they are restricted to the case where $\Phi\mathcal{D}$ is a well conditioned frame [92], or for compressed sensing [11], [33] when Φ is random. These results roughly show that in these restricted cases, $\|f^* - f^0\|^2 = O(\sigma^2)$ where σ^2 is the variance of the noise, if $\|f^0 - f_M\|$ decays fast and if μ is well chosen. This means that the noise is kept under control during the inversion.

III. TREE-STRUCTURED REPRESENTATIONS

To improve the representation of geometrical images, one can use a family of bases instead of a fixed basis. A dictionary of orthogonal bases is a set $\mathcal{D}_\Lambda = \{\mathcal{B}^\lambda\}_{\lambda \in \Lambda}$ of orthogonal bases $\mathcal{B}^\lambda = \{\psi_m^\lambda\}_m$ of $\mathbb{R}^N$, where N is the number of pixels in the image. Instead of using an a priori fixed basis such as the wavelet basis, one chooses a parameter $\lambda^* \in \Lambda$ adapted to the structures of the image to process and then uses the optimized basis $\mathcal{B}^{\lambda^*}$.

A. Quadtree-based Dictionaries

To enable the fast optimization of a parameter λ^* adapted to a given image f^0 to process, each $\lambda \in \Lambda$ is constrained to be a quadtree. The quadtree λ that parametrizes a basis $\mathcal{B}^\lambda$ defines a dyadic segmentation $[0,1]^2 = \bigcup_{(j,i) \in L(\lambda)} S_{j,i}$

where each $S_{j,i}$ is a square of size $2^{-j} \times 2^{-j}$. Figure 3 shows an example of such a dyadic subdivision. This segmentation is used to partition either the spacial domain (local cosine packets and bandlets) or the frequency domain (wavelet packets).
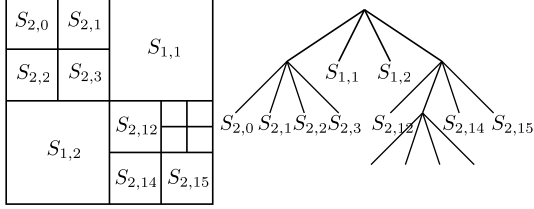

Fig. 3. *Left: example of dyadic subdivision of $[0,1]^2$ in squares $S_{j,i}$; right: corresponding quad-tree λ.*

The full tree Λ indexes sub-spaces $V_{j,i}$ of $\mathbb{R}^N$ for scales $0 \leqslant j \leqslant J$ and position $0 \leqslant i < 2^{2j}$, that obey a refinement relationship $V_{j,i} = \oplus_{\eta=0}^{3} V_{j+1,4i+\eta}$, the sum being orthogonal.

To enrich the representation parameterized by a quadtree, a token $\ell_{j,i} \in \Omega$ is attached to each leave $(j,i) \in L(\lambda)$ of the tree. We call such a tree with its token a tagged quadtree, and use the notation $\lambda \in \Lambda$ to refer to both the quadtree and its tokens.

Each sub-space $V_{j,i}$ has dimension $N/2^{2j}$ and is equipped with one or several orthogonal bases indexed by a token $\ell \in \Omega$,

$$\mathcal{B}_{j,i}^{\ell} = \left\{ \psi_{j,i,s}^{\ell} \setminus \forall\, 0 \leqslant s < N/2^{2j} \right\}.$$

Having multiple bases at each node increases the expressiveness of the dictionary, and is required to perform directional approximation of geometric images.

A basis $\mathcal{B}^\lambda$ is obtained by aggregating bases $\mathcal{B}_{j,i}^{\ell}$ for (j,i) that are leaves of λ, and for a specific choice $\ell = \ell_{j,i} \in \Omega$ of token at each node of the tree

$$\mathcal{B}^\lambda = \bigcup_{(j,i) \in L(\lambda)} \mathcal{B}_{j,i}^{\ell_{j,i}}$$

When one does not care about the location of the basis elements in the tree, the basis is written as $\mathcal{B}^\lambda = \{\psi_m^\lambda\}_m$ where the index is $m = (j,i,s,\ell_{j,i})$ with $(j,i) \in L(\lambda)$ and $0 \leqslant s < N/2^{2j}$.

This dictionary $\mathcal{D}_\Lambda$ defines a highly redundant set of atoms $\mathcal{D} = \left\{\psi_m^\lambda \setminus \lambda \in \Lambda, m\right\}$. One could use directly $\mathcal{D}$ to compute a sparse approximation using for instance ℓ^1 minimization (6). This is however numerically expensive for large images.

Restricting the sparsity to orthogonal bases leads to fast algorithms for structured dictionaries.

B. Best Basis Selection

Generalizing (4) to an adaptive setting, the approximation in a best basis is defined as $(f^\star, \lambda^\star)$ being a solution of

$$\min_{f \in \mathbb{R}^N, \lambda \in \Lambda} E(f, \lambda) = \|f^0 - f\|^2 + T^2 J^\lambda(f) \quad (10)$$

$$\text{where} \quad J^\lambda(f) = \left| \left\{ m \setminus \langle f, \psi_m^\lambda \rangle \neq 0 \right\} \right|,$$

Note that this minimization fits into the general framework (2) with $\Phi = \text{Id}$, $w = 0$, $\mathcal{E} = 0$ and $2\mu = T^2$. Note also that $f^\star$ is the best M terms approximation of f^0 with any basis in $\mathcal{D}_\Lambda$, where M is the number of non-zero coefficients of $f^\star$ in $\mathcal{B}^{\lambda^\star}$.

According to (5), for each λ, the optimal image that minimizes $E(f, \lambda)$ is $H_T(f^0, \mathcal{B}^\lambda)$. The minimization (10) can thus be simplified as $f^\star = H_T(f^0, \mathcal{B}^{\lambda^\star})$ where

$$\lambda^\star \in \operatorname*{argmin}_{\lambda \in \Lambda} \mathcal{L}_T(f^0, \mathcal{B}^\lambda) = \sum_m \gamma_T(\langle f^0, \psi_m^\lambda \rangle) \quad (11)$$

where $\gamma_T(a) = \max(a^2, T^2)$. This kind of Lagrangian is efficiently optimized using a dynamic search algorithm, originally presented by Coifman et al. [20]. It is a particular instance of the Classification and Regression Tree (CART) algorithm of Breidman et al. [7] as explained by Donoho [34].

The complexity of the algorithm is proportional to the complexity of computing the whole set of inner products $\left\{ \langle f^0, \psi_m^\lambda \rangle \setminus \lambda \in \Lambda, m \right\}$ in the dictionary. For several dictionaries, such as those considered in this section, a fast algorithm performs this computation in $O(P)$ operations where P is the total number of atoms in $\mathcal{D}_\Lambda$. For tree structured dictionaries, this complexity is thus $O(|\Omega|N \log_2(N))$, where $|\Omega|$ is the number of tokens associated to each leaf of the tree. This is much smaller than the total number of bases $\mathcal{B}^\lambda$ in $\mathcal{D}_\Lambda$, which grows exponentially with N.

C. Spacial and Frequency Adaptive Dictionaries

Local cosine and wavelet packets bases $\mathcal{B}^\lambda$ are parameterized by a quadtree tree $\lambda \in \Lambda$ that segments respectively the image plane and the frequency plane [22], [61], [20].

The wavelet packet construction is generalized by considering non stationary (NS) wavelet packets [18], that apply different quadrature mirror filters at each scale of the tree. A NS wavelet packet basis $\mathcal{B}^\lambda$ is parameterized by a tagged quadtree λ where each node is tagged by a filter. The selection of a best NS wavelet packet basis is obtained by a dynamic programming algorithm detailed in [71] that generalizes the best basis algorithm.

D. Geometry Adaptive Dictionaries

Geometric images with regular edges are difficult to approximate using a wavelet basis. Wedglets approximations and bandlets bases compute a quadtree segmentation $\lambda^\star$ of the image together with a local estimate of the orientation of the edges, that is encoded in the tokens $\ell_{j,i} \in \Omega$.

Adaptive wedgelets approximation. A geometric approximation is obtained by considering for each node (j,i) a collection of discontinuous approximation spaces for the square $S_{j,i}$ of the segmentation. The wedgelets [32] uses a piecewise constant approximation on each side of a straight edge that follows the direction indexed by $\ell_{j,i} \in \Omega$.

The best wedgelets algorithm optimizes the Lagrangian (11) over all possible wedgelets segmentation $\lambda \in \Lambda$ where the thresholding approximation is replaced by a piecewise constant approximation. This scheme is efficient to approximate a piecewise constant image f^0 whose edges are C^2 curves. For such a cartoon image, the approximation error with M wedgelets decays like $\|f^0 - f_M\|^2 = O(M^{-2})$, see [47]. It is also possible to consider approximation spaces with higher order polynomials to capture arbitrary cartoon images [88]. The computation of the low-dimensional projection can be significantly accelerated, see [46].

Adaptive spatial Bandlet approximation. The bandlet bases dictionary was introduced by Le Pennec and Mallat [55], [56]. Bandlets perform an efficient adaptive approximation of images with geometric singularities. This transform has been refined by Mallat and Peyré [62], [79] to obtain a dictionary of regular and orthogonal basis functions.

A directional orthogonal basis $\mathcal{B}^\ell_{j,i}$ of elongated directional wavelets is defined over each square $S_{j,i}$. Keeping only a few bandlet coefficients and setting the others to zero performs an approxima-

tion of the original image that follows the local direction indexed by $\ell_{j,i}$. The optimal segmentation and tokens are those that minimize the Lagrangian (11). Figure 4 shows an example of such a segmentation adapted to a geometric image.
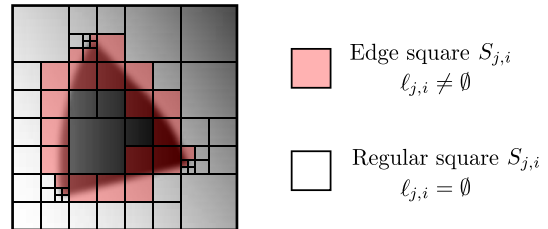

Fig. 4. *Example of subdivision $\lambda^\star$ adapted to a cartoon image f^0. The basis $\mathcal{B}^{\lambda^\star}$ minimizes $\mathcal{L}_T(f^0, \mathcal{B}^\lambda)$.*

Adaptive approximation over the wavelet domain. Applying such an adaptive geometric approximation directly on the image leads to unpleasant visual artifacts. To overcome this issue, one applies a tree structured approximation on the discrete set of wavelet coefficients. The wedgeprints of Wakin et al. [93] use a vector quantization to extend the wedgelet scheme to the wavelet domain. The orthogonal bandlets of Peyré and Mallat [62] use an adaptive bandlet basis for each scale of the wavelet transform. The orthogonal bases are computed using discrete oriented orthogonal bases. If f^0 is a function that is C^α outside a set of C^α-regular edge curves, the approximation in the best bandlet basis $\mathcal{B}^{\lambda^\star}$ minimizing (11) satisfies $\|f^0 - f_M\|^2 = O(M^{-\alpha})$ where M is the number of bandlets composing $f_M = H_T(f^0, \mathcal{B}^{\lambda^\star})$. Such an orthogonal bandlet transform finds applications to image [62] and surface [79] compression and image denoising [80].

E. Adaptive Tree-structured Processing

The processing scenarios described in Section II-C extend to the best basis setting. While compression and denoising only require the estimation of a best basis parameter $\lambda^\star$ as a pre-processing, solving inverse problems requires a more complicated iterative scheme.

Adaptive tree-structured compression. For compression applications, one computes the best basis $\mathcal{B}^{\lambda^\star}$ adapted to the image f^0 to compress by minimizing the corresponding Lagrangian (11), and then binary codes the quantized coefficients

$\langle f^0, \psi_m^\lambda \rangle$ together with the parameter λ of the best basis, see [61].

The coding error decay (7)extends to a dictionary of orthognonal basis if $|\mathcal{D}| = P = O(N^\beta)$ for some $\beta > 0$, see [61]. The total number of bits R is reduced by splitting it as $R = R_c + R_{\lambda^\star}$ where R_c is the number of bits to code the M non-zero coefficients among N coefficients, and $R_{\lambda^\star}$ is the number of bits to code the quadtree $\lambda^\star$.

For compressing C^2 cartoon images, the best wedgelet sub-space provides a distortion rate $\|f^0 - f_R\|^2 = O(R^{-2} \log^2(R))$. The best bandlet basis further improve this rate to $\|f^0 - f_R\|^2 = O(R^{-\alpha} \log^\alpha(R))$ for cartoon images that are C^α outside C^α edges.

Adaptive tree-structured denoising. To perform denoising, the best basis $\mathcal{B}^{\lambda^\star}$ is computed from the noisy observations $y = f^0 + w$ by minimizing an empirical Lagrangian $\mathcal{L}_T(y, \mathcal{B}^\lambda)$, see [61], [54]. The denoised image is defined by thresholding the noisy observations $f^\star = H_T(y, \mathcal{B}^{\lambda^\star})$. The asymptotic optimality result (8) extends to a dictionary of orthogonal bases containing $|\mathcal{D}| = P$ atoms by selecting $T = \sigma\sqrt{2\log(P)}$, [61].

best sub-space was originally designed to perform denoising.

For piecewise constant function with C^2 edges, Wedgelets leads to an optimal decay of the estimation error of $E_w(\|f^0 - f^\star\|)^2 = O(\sigma^{4/3})$, up to logarithmic factor, where σ^2 is the variance of the noise [32], [30]. The best bandlet basis thresholding extends this result to a decay of $E_w(\|f^0 - f^\star\|)^2 = O(\sigma^{\frac{2\alpha}{\alpha+1}})$ for C^α cartoon images, see [80].

Adaptive Tree-structured Regularization of Inverse Problems The ℓ^1 resolution (9) of the inverse problem $y = \Phi f^0 + w$ is extended to a best basis recovery by solving over both the image and the quadtree

$$\min_{f \in \mathbb{R}^N, \lambda \in \Lambda} E(f, \lambda) = \frac{1}{2}\|y - \Phi f\|^2 + \mu J^\lambda(f) \quad (12)$$

$$\text{where} \quad J^\lambda(f) = \sum_{m=0}^{N-1} |\langle f, \psi_m^\lambda \rangle|,$$

which corresponds to the general framework introduced in (2)with $\mathcal{E} = 0$.

Peyré proposes in [75] an iterative algorithm to compute a stationary point of E. Starting from some initial $f^{(0)}$, at a step k, the estimate $f^{(k)}$ is modified as

$$\tilde{f}^{(k)} = f^{(k)} + \tau\Phi^*(y - \Phi f^{(k)}).$$

The best basis parameter is then updated

$$\lambda^{(k+1)} = \arg\min_{\lambda \in \Lambda} \sum_{m=0}^{N-1} \gamma_\mu^1(\langle \tilde{f}^{(k)}, \psi_m^\lambda \rangle)$$

$$\text{where} \quad \gamma_\mu^1(a) = \begin{cases} a^2/2 & \text{if} \ |a| \leqslant \mu, \\ \mu|a| - \mu^2/2 & \text{otherwise.} \end{cases}$$

Note that this is a Lagrangian optimization similar to (11), which can be solved efficiently using the same dynamic programming algorithm [20]. The next iterate is defined using a soft thresholding

$$f^{(k+1)} = \sum_m s_{\tau\mu}(\langle f, \psi_m^\lambda \rangle)\psi_m^\lambda$$

$$\text{where} \quad s_{\tau\mu}(x) = \max(1 - \tau\mu/|x|, 0)x.$$

If $\tau < 1/\|\Phi^*\Phi\|$, one can show that $E(f^{(k)}, \lambda^{(k)})$ is decaying through the iterations, and thus converging, see [75].

Figure 5 shows an example of application of this algorithm to the recovery from compressed sensing measurements [11], [33] using a dictionary of bandlet basis. The operator Φ projects an image of N pixels on $P = N/6$ random vectors.

IV. SEMI-STRUCTURED REPRESENTATIONS

A. Lifting Based Dictionaries

To enhance the wavelet representation, it is possible to adapt the wavelet filters to the image content, using spatially adaptive filters to avoid creating large wavelet coefficients near edges. The lifting scheme, introduced by Sweldens [90], is an unifying framework to design adaptive biorthogonal wavelets, through the use of spatially varying local interpolations.

Lifting scheme. The lifting scheme has been introduced by Sweldens [90], and uses a polyphase decomposition of the wavelet filtering to decompose it into elementary lifting steps. This allows one to define wavelets on non-translation invariant grids using spatially varying filters. It can thus be used to define wavelets on surfaces [86], [57] and on complicated planar triangulations, see Section IV-C.

The initial sampling grid $\mathcal{V}_J$ for some $J > 0$ is recursively split, for each $j < J$ as a disjoint union $\mathcal{V}_{j+1} = \mathcal{V}_j \cup \mathcal{C}_j$, where $\mathcal{V}_j$ is a coarser grid, and $\mathcal{C}_j$ is a detail grid intended to store wavelet
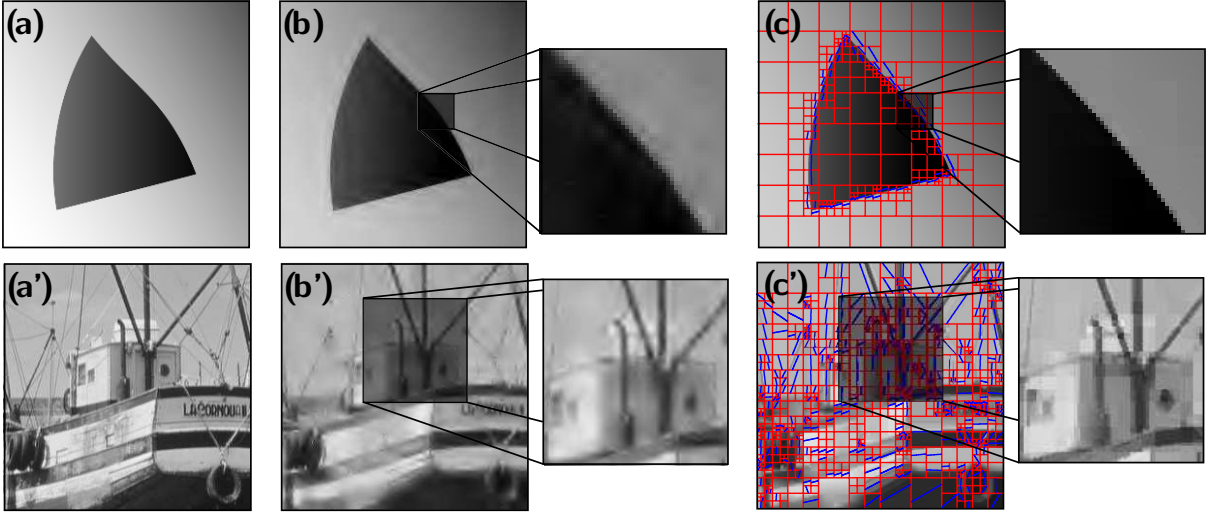
Fig. 5. *(a,a') original image ; (b,b') compressed sensing reconstruction using a translation invariant wavelet frame (PSNR=37.1dB and 22.1dB) ; (c,c') reconstruction using iteration in a best bandlet basis (PSNR=39.3dB and 23.9dB).*

coefficients. On an image, one can use a traditional dyadic refinement so that $|\mathcal{V}_j| = |\mathcal{V}_{j+1}|/4$, or to build non-tensor product wavelets, one can use quincunx sub-sampling [44] where $|\mathcal{V}_j| = |\mathcal{V}_{j+1}|/2$. But this splitting scheme is general and holds for instance in the case of triangulation refinement, see Section IV-C.

A biorthogonal wavelet basis $\mathcal{B}^\lambda = \{\psi_{j,n}^\lambda\}_{j,n \in \mathcal{C}_j}$ is implicitly defined through lifting by computing the wavelet coefficients $d_j^\lambda[n] = \langle f^0, \psi_{j,n}^\lambda \rangle$ for $n \in \mathcal{C}_j$. though the application of elementary lifting steps. The lifting algorithm also maintains temporary variables, which are inner products with scaling functions $a_j^\lambda[p] = \langle f^0, \varphi_{j,p}^\lambda \rangle$ for $p \in \mathcal{V}_j$. At the finest scale $j = J$, the scaling coefficients are initialized using the data $a_J = f$. Then, for each $j < J$, it successively applies a local predictor to compute the wavelet coefficients, for all $n \in \mathcal{C}_j$,

$$d_j^\lambda[n] = a_{j+1}^\lambda[n] - \sum_{p \in \mathcal{V}_j, p \in \lambda_j[n]} w_j[n,p] a_{j+1}^\lambda[p],$$
(13)

and an update operator to maintain stability of the transform, for all $p \in \mathcal{V}_j$,

$$a_j^\lambda[p] = a_{j+1}^\lambda[p] + \sum_{n \in \mathcal{C}_j, p \in \lambda_j[n]} \tilde{w}_j[n,p] d_j^\lambda[n]. \quad (14)$$

The association field λ_j connects each point $n \in \mathcal{C}_j$ of the detail grid to a set of points $\lambda_j[n] \subset \mathcal{V}_j$ on the coarse grid. This association field is the only degree of freedom of the scheme,

since the weights $w_j[n,p]$ are usually computed as a polynomial interpolation on the set of points selected by the field, while $\tilde{w}_j[n,p]$ is obtained by solving a linear system to impose that a_j^λ has the same mean and higher order moments as a_{j+1}^λ. The lifting wavelet transform is easily inverted by successively inverting steps (14) and (13).

One can chain more than one predict and one update operators, to construct arbitrary biorthogonal wavelets on uniform 1-D grid [26]. Unfortunately, on non-uniform grids, it is impossible to control properties such as vanishing moments, orthogonality and stability of more than one step of predict and update.

Lifting processing. A non-linear approximation is obtained by thresholding the lifted wavelet coefficients $d_j^\lambda[n]$ before reconstruction. This can be used to perform compression and denoising following the schemes detailed in Section II-C. Note however that for arbitrary association fields $\{\lambda_j\}_j$ the lifted wavelets can be unstable and far from orthogonality. This results in a loss of performance for compression and denoising using thresholding.

Adaptive predictions. One can adaptively design the association fields λ_j used during the lifting pyramid. One can compute these associations to reduce the length of the wavelet filter near edges [17]. For compression applications, this estimation is performed from already coded coefficients, so that the fields λ_j do not need to

be transmitted to the decoder. Such a filter adaptation is however no sufficient to capture fully the geometric information. Such schemes are related to adapted subdivision [65].

To further reduce the distortion of geometric images, the orientation of the association fields $\{\lambda_j\}_j$ can be optimized though the scales. Because of the lack of structure of the set of bases $\mathcal{B}^\lambda$, computing the fields $\lambda^\star$ that minimizes (10) is intractable. The lifting operators are thus usually computed using heuristics to detect the local orientation of edges. The association fields λ_j are obtained by minimizing the amplitude of wavelet coefficients, see for instance [49], [14] or using local gradient information [52]. These adaptive lifting schemes can be extended to perform adaptive video transform [87] where the lifting operates through in time by following the optical flow.

B. Grouplets

A difficulty with lifted transforms is that they do not guarantee the orthogonality of the resulting wavelet frame. The stability of the transform thus tends to degrade for highly adaptive association fields, where some singular point $p \in \mathcal{V}_j$ might belong to many different fields $\lambda_j[n]$ for various $n \in \mathcal{C}_j$.

Grouplets over the image. The grouplet transform also makes use of an association field, but it replaces the lifting computation of wavelet coefficients by an extended Haar transform, where points in $\mathcal{C}_j$ are processed in sequential order to maintain orthogonality.

A grouplet family $\mathcal{B}^\lambda$ is an orthogonal basis, or a redundant tight frame adapted to the processing of geometric images or regularly oscillating textures, introduced by Mallat [60]. It has been used to perform image denoising and super-resolution [60] as well as texture inpainting and synthesis [76]. The geometric parameter $\lambda = \{\lambda_j\}_j$ is a multiscale association field, where each λ_j is a vector field that follows the geometry of the image at a scale 2^j.

For each scale j, the flow λ_j links the point x to $\lambda_j(x)$, where $\|x - \lambda_j(x)\| \approx 2^j$. A modified Haar transform progressively extract grouplets detail coefficients along this flow. This corresponds to the projection of the image on a orthogonal or redundant tight frame $\mathcal{B}^\lambda = \{\psi_{j,n}\}_{j,n}$ of $\mathbb{R}^N$. The grouplets atoms at scale j have a width of 1 pixel and follow the geometric flow on a width of $\sim 2^j$

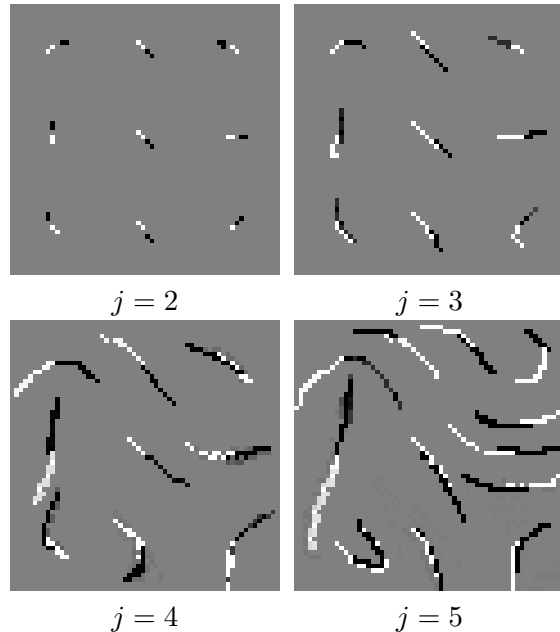pixels. Image 6 shows examples of grouplets at several scales.

$j = 2$ $j = 3$

$j = 4$ $j = 5$

Fig. 6. *Examples of grouplet vectors $\psi_{j,n}$ for several positions n and scale j.*

The selection of an optimal grouplet flow $\lambda^\star$ adapted to an image f^0 by minimizing a Lagrangian such as (11) is intractable, because the grouplet dictionary does not have a tree structure. One thus uses a greedy scheme that performs block matching to select the patch around $\lambda_j(x)$ that is the closest to the patch around x. Alternative methods compute λ_j by integrating a texture flow along a distance of 2^j, see [76].

The grouplet transform is related to the curved wavelet transform [94] and the easy path [81]. A distinctive features of grouplets is that the fields λ can converge to a singular point, which means $\lambda_j(x) = \lambda_j(x')$ for $x \neq x'$. This is useful to represent turbulent textures with complicated singularities.

Grouplets over wavelets. To process image with both oriented texture and edges, this grouplet transform can also be applied over a wavelet transform, so that the wavelet coefficients at each wavelet scale are projected on a grouplet frame $\mathcal{B}^\lambda$. This corresponds to the projection of f^0 on a multiscale grouplet frame composed of atoms with varying widths and lengths.

This grouplet transform has been applied to texture synthesis [76], see Figure 7. The synthe-

sis is obtained by sampling grouplet coefficients with the same distribution as the coefficient of an exemplar, and by randomizing the geometry of the flow λ.
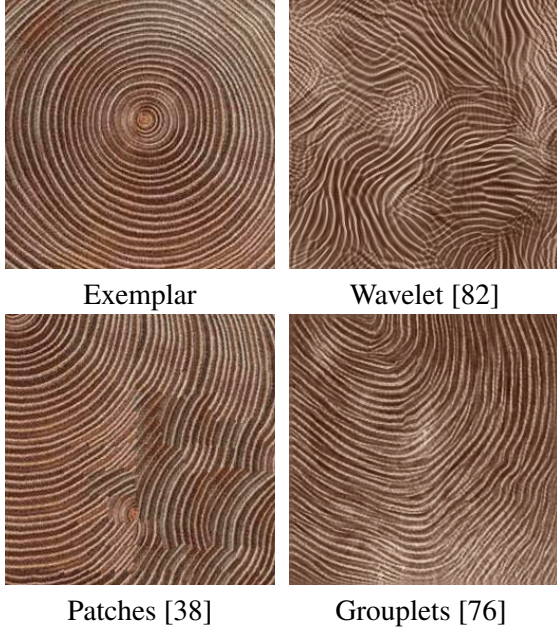

Exemplar   Wavelet [82]

Patches [38]   Grouplets [76]

Fig. 7. *Example of texture synthesis by statistical modeling of grouplet coefficients.*

Fig. 8. *A few iterations of the grouplet inpainting algorithm*

Sparse grouplet regularization can be used to perform geometric inpainting of large missing regions by minimizing (12) over the image f^0 while adapting progressively the association field λ, see [76]. Figure 8 shows an example of iterations of the grouplet inpainting algorithm, that performs iteratively a soft thresholding and an estimation of an adapted grouplet flow λ.

C. Adaptive Triangulations

A finite elements approximation first builds a segmentation of the image domain into elementary cells, and then perform a low dimensional polynomial approximation on each cell. The adaptivity comes from the selection of this segmentation, that should be adapted to the features of the image.

Image approximation by triangulations. The parameters $\lambda = (\mathcal{V}, \mathcal{F})$ to adapt a triangulation dictionary is a set of vertices $\mathcal{V} = \{x_m\}_{m=0}^{M-1} \subset [0,1]^2$ and a set of connexions $\mathcal{T} \subset \mathcal{V}^3$ that prescribes the connectivity between the points. Triangles are convex hulls of the elements of $\mathcal{T}$, and they form a partition of the image domain $[0,1]^2$.
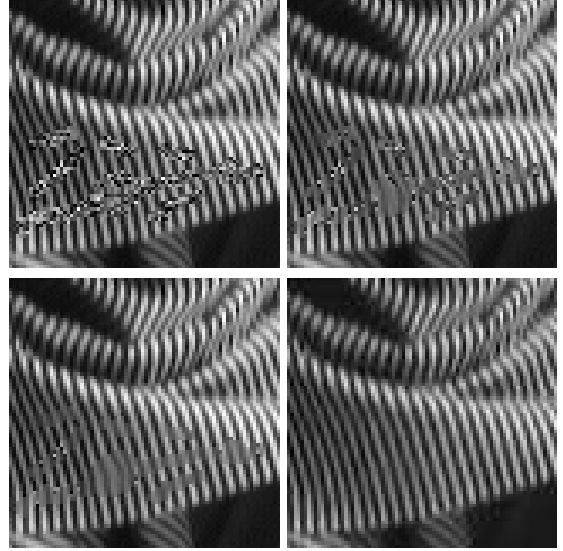
One performs approximation with continuous piecewise affine function on the triangles, which corresponds to using an adapted basis $\mathcal{B}^\lambda = \{\psi_m\}_{x_m \in \mathcal{V}}$ of the space V_λ of linear spline function (hat functions centered on the vertices $x_m \in \mathcal{V}$), which has dimension M.

The best approximation with M vertices of f^0 is obtained by an orthogonal projection

$$f_M = P_{V_\lambda} f^0 = \sum_{m=0}^{M-1} a_m^\star \psi_m \qquad (15)$$

where $\quad a^\star = \underset{a \in \mathbb{R}^M}{\operatorname{argmin}} \|f^0 - \sum_m a_m \psi_m\|^2.$

Since $\mathcal{B}^\lambda$ is not orthogonal, computing the coefficients a requires the resolution of a sparse $M \times M$ linear system.

Smooth image approximation with triangulations. For $\mathcal{C}^2$ uniformly regular functions, one can build isotropic triangulations that generate an error $\|f^0 - f_M\|^2 = O(\|H_{f^0}\|_{L^2}^2 M^{-2})$, where H_{f^0} is the Hessian of the image f^0.

For a C^2 image, one should use anisotropic triangles whose aspect ratio matches the anisotropy of the image [4]. A triangle near a point $x \in [0,1]^2$ should be aligned with the principal eigenvector of $H_{f^0}(x)$, and the aspect ratio $width/length$ should match the ratio of the eigenvalues of $H_{f^0}(x)$. The approximation error with such adapted anisotropic approximations of a C^2 image satisfies $\|f^0 -$

$f_M\|^2 = O(\|\sqrt{|\det(H_{f^0})|}\|^2_{L^{3/2}}M^{-2})$, see for instance [69]. This shows that anisotropic approximations have a better constant than isotropic for the decay of the error $\|f^0 - f_M\|^2$.

Cartoon image approximation with triangulations. For a cartoon image, which is a $\mathcal{C}^2$ function outside $\mathcal{C}^2$ contours, one can also diminish the approximation error $\|f^0 - f_M\|^2$ using anisotropic triangles. Near edges, the anisotropic triangles should obey the same parabolic scaling as curvelet atoms, $width = length^2$. Figure 9 shows an example of such an adapted triangulation. Using such anisotropic triangulations for a cartoon image f^0 leads to an error decay $\|f^0 - f_M\|^2 = O(M^{-2})$, see [61]. This improves significantly over the wavelet approximation decay $O(M^{-1})$.
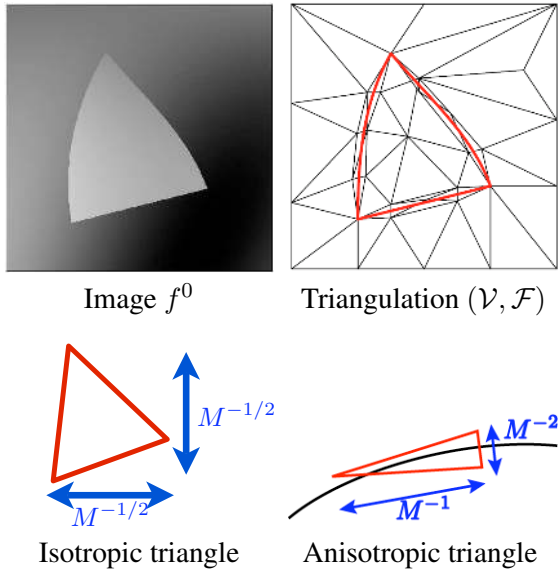

Fig. 9. *Adaptive triangulations for piecewise linear approximations of a $\mathcal{C}^2$-cartoon image.*

Heuristics and greedy algorithms. The issue is that computing the best triangulation $(\mathcal{V}, \mathcal{F})$ that minimizes $\|f^0 - f_M\|$ over a discretized domain is NP-hard [1]. There is currently no algorithm that guarantees the error decay $\|f^0 - f_M\|^2 = O(M^{-2})$ for an arbitrary cartoon image. One thus needs to rely on sub-obtimal greedy schemes. These schemes generate a sequence of embedded triangulation $\{\lambda_j = (\mathcal{V}_j, \mathcal{F}_j)\}_j$ by either refinement (increasing j values) or coarsening.

A greedy refinement scheme starts by a simple fixed triangulation λ_0 of the squares $[0,1]^2$, and iteratively add one or several vertices to $\mathcal{V}_j$ to
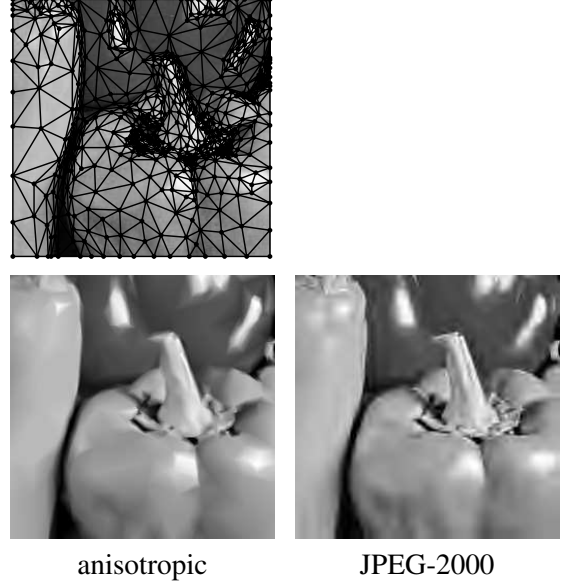

Fig. 10. Comparison of the adapted triangulation scheme [6] with JPEG-2000, for the same number of bits, $M = 600$ triangles.

obtain a triangulation λ_{j+1} that minimizes the error $\|f^0 - P_{V_{\lambda_j}} f^0\|^2$. The faces $\mathcal{F}_j$ needs also to be updated to obtain $\mathcal{F}_{j+1}$.

Delaunay refinement introduced by Ruppert [85] and Chew [16], proceeds by inserting a single point, that is imposed to be a circumcenter of one triangle, and also impose that $\mathcal{F}_j$ is a Delaunay triangulation of $\mathcal{V}_j$. This constraint accelerates the search and also leads to triangles with provably good isotropic aspect ratio, which might be useful to compute the approximation of the solution of an elliptic PDE on the mesh grid. For image approximation, one however needs to design aniosotropic triangulations, which requires to modify the notion of circumcenter using an anisotropic metric [48], [6]. Figure 10 shows an example of adaptive triangulation produced by the greedy anisotropic refinement scheme of [6]. Other refinement schemes are possible, such as for instance edge bisection, for which the optimal error decay is known to hold for smooth convex functions [69].

Triangulation thinning algorithms start with a fine scale triangulation λ_J of $[0,1]^2$ and progressively remove either a vertex, an edge or a face to increase as slowly as possible the approximation error $\|f^0 - P_{V_{\lambda_j}} f^0\|^2$ until M vertices remain, see for instance [36]. One can for instance remove a

single vertex to go from $\mathcal{V}_{j+1}$ to $\mathcal{V}_j$, and impose that $\mathcal{F}_j$ is the Delaunay triangulation of $\mathcal{V}_j$. This can be shown experimentally to produce highly anisotropic meshes, which can be used to perform compression, see Demaret et al. [29].

Multiscale approximations on triangulations. Embedded triangulations $\lambda_j = (\mathcal{V}_j, \mathcal{F}_j)$ with an increasing number of vertices satisfy $V_{\lambda_j} \subset V_{\lambda_{j+1}}$. This is achieved by imposing that $\mathcal{V}_j \subset \mathcal{V}_{j+1}$ and that each face of $\mathcal{F}_j$ is split into one of more faces of $\mathcal{F}_{j+1}$. The set of triangulation $\lambda = \{\lambda_j\}_j$ defines a hierarchical segmentation of the image that can be encoded using a tree, generalizing the quadtree construction of Section III-A. A biorthogonal wavelet basis $\mathcal{B}^\lambda = \{\psi_{j,n}\}_{n \in \mathcal{C}_j}$ is constructed using a basis $\{\psi_{j,n}\}_{n \in \mathcal{C}_j}$ of the detail space W_{λ_j} that satisfies $V_{\lambda_{j+1}} = V_{\lambda_j} \oplus W_{\lambda_j}$. The decomposition of an image on a biorthogonal wavelet basis defined on embedded triangulations can be computed using lifting steps, see [86], [57].

Regular splits of orthogonal triangles lead to isotropic adaptive triangulations [31]. Splitting triangles according to a well chosen median leads to anisotropic triangulations that exhibit optimal aspect ratio for smooth images, see [19].

V. Patch-based Representations

A new class of adaptive image processing have emerged recently by considering non-local interactions between patches in an image. These interactions are used to perform non-local filtering, sparse coding with dictionary learning, and non-local regularization of inverse problems.

A. Manifold Geometry of Patches

Manifold of patches. A patch $\pi_x(f)$ of size $\tau \times \tau$ is extracted around a pixel x in an image $f \in \mathbb{R}^N$ by defining for all $y \in \{-\tau/2 + 1, \ldots, \tau/2\}^2$,

$$\pi_x(f)(y) = f(x + y). \quad (16)$$

The set of patches $\{\pi_x(f)\}_x$ extracted from natural images often depends on a small number of hidden variables. This is for instance the case near geometric edges or directional textures.

For some simple classes of geometric images and textures, this set of patches can be shown to be close to a low dimensional smooth manifold $\lambda \subset \mathbb{R}^{\tau^2}$ of dimension $R \ll \tau^2$. This is for instance the case for binary cartoon images, where edge patches can be parameterized by the orientation

of the closest edge, and the distance to this edge. The resulting manifold λ has the topology of a cylinder, as shown in Figure 11. Similar manifolds can be use to model more general cartoon images and oriented textures, see [73].
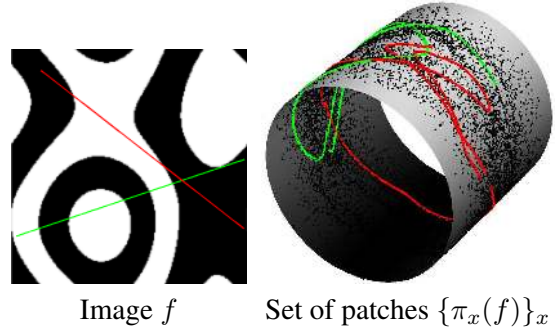

Image f Set of patches $\{\pi_x(f)\}_x$

Fig. 11. *Display of the patch manifold that has the topology of a cylinder for binary edges.*

Manifold regularization. The manifold model assumes that the patches $\pi_x(f)$ are close to the manifold λ, and a variational energy quantifies the deviation from this model

$$J^\lambda(f) = \frac{1}{2} \sum_x d(\pi_x(f), \lambda)^2 \quad \text{where}$$

$$\forall p \in \mathbb{R}^{\tau^2}, \quad d(p, \lambda)^2 = \min_{q \in \lambda} \|p - q\|^2.$$

For simple geometric images and textures, where λ is explicitly known, the energy J^λ is a fixed, non adaptive prior. It can be used to regularize an inverse problem $y = \Phi f^0 + w$ by minimizing

$$\min_{f \in \mathbb{R}^N} E(f, \lambda) = \frac{1}{2} \|y - \Phi f\|^2 + \mu J^\lambda(f) \quad (17)$$

$$= \min_{f \in \mathbb{R}^N, \{\pi_x\}_x \in \lambda^N} \|y - \Phi f\|^2 + \mu \sum_x \|\pi_x(f) - \pi_x\|^2.$$

A stationary point of this non-convex energy E is computed using an alternative minimization on f and on the patch $\{\pi_x\}_x \in \lambda^N$ of the manifold, see [73]. Figure 12 shows an example of compressed sensing reconstruction using a regularization with a manifold of edges. If the manifold λ is also optimized in (17), it corresponds to an adaptive regularization that fits into the framework (2).

Geometric texture regularization. Separating a texture content f_1 from the cartoon content f_2 of a noisy image $y = f_1 + f_2 + w$ has been emphasized
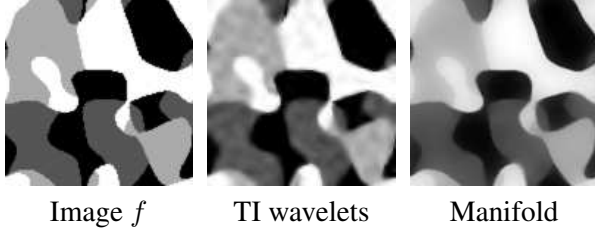
Image f TI wavelets   Manifold

Fig. 12. *Example of recovery from $P = N/6$ compressed sensing measurements, using sparsity in a translation invariant wavelet frame, and using (17) for the manifold of edges.*

by Yves Meyer [67] that introduces several norms for oscillating patterns. Several numerical algorithms have proposed to perform this separation, see for instance [3].

Locally parallel textures are oscillating in a direction indicated at a point x by a smooth vector field $\lambda(x)$. An adaptive texture norm makes use of the regularity of the patterns and enhances the separation with respect to non-adaptive norms that favor arbitrary oscillations. Figure 13 shows the local Fourier expansion of an image f mixing edges and textures. This local analysis is obtained by projecting f on a local Fourier frame $\{\psi_{p,k}\}_{p,k}$. Each atom $\psi_{p,k}$ is localized in a patch π_{x_k} around a point x_k and has a frequency ξ_k.
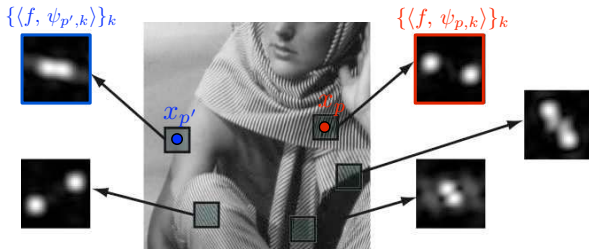

Fig. 13. *Local Fourier analysis of an image containing edges and textures.*

Following Maurel et al. [66], an adaptive texture energy $J^\lambda(f)$ is defined as a weighted quadratic sum of the local Fourier coefficients

$$\frac{1}{2} \sum_{p,k} \gamma(\|\xi_k - \lambda(x_p)\|)\gamma(\|\xi_k + \lambda(x_p)\|)|\langle f, \psi_{p,k}\rangle|^2$$

where $\gamma(r)$ is a weighing function that is close to zero when $r \approx 0$ and is rapidly increasing with r. This weighting function forces the local frequencies of the texture to be non-zero around x_p

only for frequencies close to $\pm\lambda(x_p)$. This prior is well suited to model locally parallel textures.

The energy J^λ captures the oscillating content of the image. The cartoon content of the image can be captured using the total variation $J_{TV}(f) = \sum_x \|\nabla f(x)\|$, where $\nabla f(x) \in \mathbb{R}^2$ is a finite difference approximation of the gradient of f at x. The cartoon and the texture priors are integrated into a variational energy minimization to perform a separation $f_1^\star + f_2^\star$ from noisy low dimensional measurement $y = \Phi f^0 + w$

$$(f_1^\star, f_2^\star, \lambda^\star) \in \operatorname*{argmin}_{f_1,f_2,\lambda} E(f_1, f_2, \lambda) \qquad (18)$$

$$= \frac{1}{2}\|y - \Phi(f_1 + f_2)\|^2 + \mu J^\lambda(f_2) + \nu J_{TV}(f_1) + \mathcal{E}(\lambda),$$

where μ, ν should be adapted to the noise level and where $\mathcal{E}$ can be used to encode constraints on the frequency localization of the texture. This separation problem requires the computation of both the components and the adapted geometric flow λ. It generalizes the adaptive regularization (2) to the case where several priors are used. An iterative minimization algorithm computes a stationary point of E, see [66]. Figure 14 shows an example of decomposition obtained by solving (18), using $\Phi = \mathrm{Id}$ the identity operator.
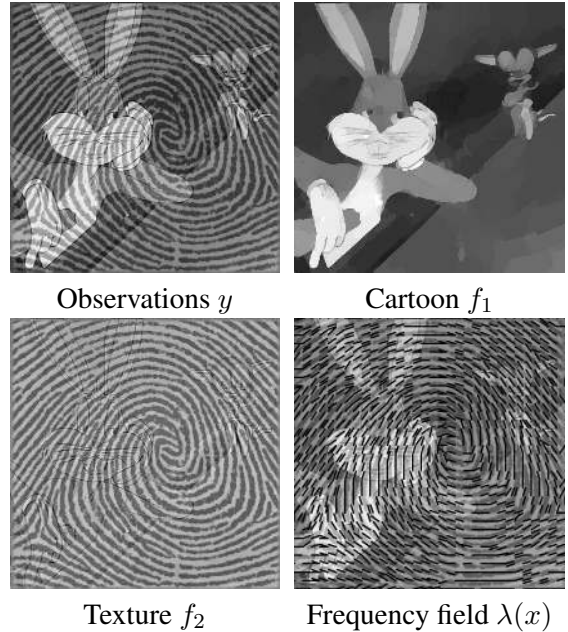

Observations y Cartoon f_1

Texture f_2 Frequency field $\lambda(x)$

Fig. 14. *Variational separation using an energy adapted to locally parallel textures.*

B. Dictionary Learning

Dictionary learning for patch approximation.
Sparse modeling of the patches $\{\pi_x\}_x$ requires the design of a dictionary $\mathcal{D} = \{\psi_m\}_{m=0}^{P-1}$ adapted to the local geometry of the images to process. The dictionary $\mathcal{D}$ can be learned from a large set of $Q \gg P$ patches $\{\pi_k\}_{0 \leqslant k < Q}$. As first proposed by Olshausen and Field [70], $\mathcal{D}$ is optimized to maximize the sparsity of these exemplars, see also [42], [2] and references therein.

The optimal dictionary minimizes the average approximation error using the basis pursuit sparse coding (6)

$$\min_{\mathcal{D}=\{\psi_m\}_m, \{a_k\}_k} \sum_{k=0}^{Q-1} \left\| \pi_k - \sum_{m=0}^{P-1} a_{k,m}\psi_m \right\|^2 \quad (19)$$

$$+\mu \sum_{k=0}^{Q-1} \sum_{m=0}^{P-1} |a_{k,m}| + \mathcal{E}(\mathcal{D})$$

and where the constraint on the dictionary is defined as

$$\mathcal{E}(\mathcal{D}) = \begin{cases} 0 & \text{if} \quad \forall m, \; \|\psi_m\| \leqslant 1, \\ +\infty & \text{otherwise.} \end{cases} \quad (20)$$

This problem is convex over the dictionary (under convex constraints) and on the coefficients, but not jointly convex. One can use an iterative coordinate descent that converges to a stationary point of this minimization, see for instance [78]. When the dictionary $\{\psi_m\}_m$ is fixed, one perform a sparse coding to find the coefficients that solve, for each k the basis pursuit denoising (6)

$$\min_{a_k \in \mathbb{R}^P} \left\| \pi_k - \sum_{m=0}^{P-1} a_{k,m}\psi_m \right\|^2 + \mu \sum_{m=0}^{P-1} |a_{k,m}|. \quad (21)$$

When the coefficients $\{a_{k,m}\}_{k,m}$ are fixed, the dictionary is the solution of

$$\min_{\{\psi_m\}_m, \|\psi_m\| \leqslant 1} \sum_{k=0}^{Q-1} \left\| \pi_k - \sum_m a_{k,m}\psi_m \right\|^2$$

which is the minimization of a quadratic convex functional under quadratic convex constraints, and can be solved using a projected gradient descent [78].

Figure 15 shows an example of dictionary learned with this algorithm. The exemplars $\{\pi_k\}_k$ are patches of 16×16 pixels extracted from a library of natural images, and the redundancy of the dictionary is set to $P/N = 2$.
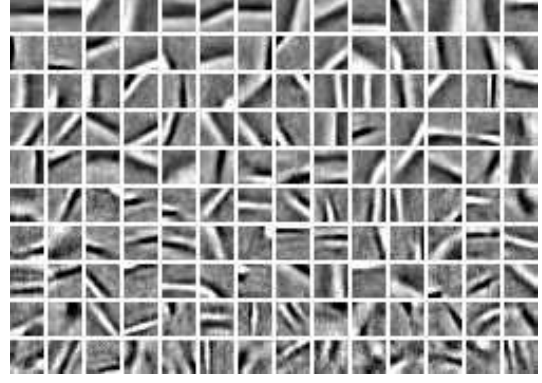

Fig. 15. *Some vectors ψ_m of a dictionary $\mathcal{D}$ learned from a set of patches extracted from natural images.*

Application to denoising and inverse problems.
This learning can be performed jointly with the resolution of an inverse problem $y = \Phi f^0 + w$ by minimizing

$$\min_{f \in \mathbb{R}^N, \mathcal{D}=\{\psi_m\}_m} \frac{1}{2}\|y - \Phi f\|^2 + \mu J^{\mathcal{D}}(f) + \mathcal{E}(\mathcal{D}), \quad (22)$$

where $\mathcal{E}$ is defined in (20). The prior $J^{\mathcal{D}}$ enforces the sparsity in $\mathcal{D}$ of each patch $\pi_{x_k}(f)$ extracted at some location x_k as defined in (16)

$$J^{\mathcal{D}}(f) = \sum_{k=0}^{Q-1} \min_{a_k \in \mathbb{R}^P} \frac{1}{2}\left\| \pi_{x_k}(f) - \sum_{m=0}^{P-1} a_{k,m}\psi_m \right\|^2 \quad (23)$$

$$+\eta \sum_{m=0}^{P-1} |a_{k,m}|.$$

Note that this corresponds to the adaptive regularization framework (2) with the replacement of the geometry λ by a patch dictionary $\mathcal{D}$.

The resolution of (22) requires the minimization of a non-convex energy over both the image to recover and the dictionary to learn. This minimization can be performed alternatively on f, the coefficients a_k involved in (23), and $\{\psi_m\}_m$, which are all convex problems. One can show the convergence of this scheme toward a stationary point of the energy, see [78].

An indepth review of the applications of dictionary learning to image processing can be found in [40]. For $\Phi = \text{Id}$, the minimization (22) performs state of the art denoising [39]. When Φ is a masking operator, the minimization (22) performs inpainting [59], in which case the dictionary can be learned from exemplar images, and also updated

at each step of the minimization. One can also use several dictionaries and perform a joint image separation and regularization. This separation process extends the original morphological component analysis to an adaptive setting [78]. Figure 16 shows the enhancement of the inpainting obtained using a learned dictionary instead of a fixed local cosine dictionary.
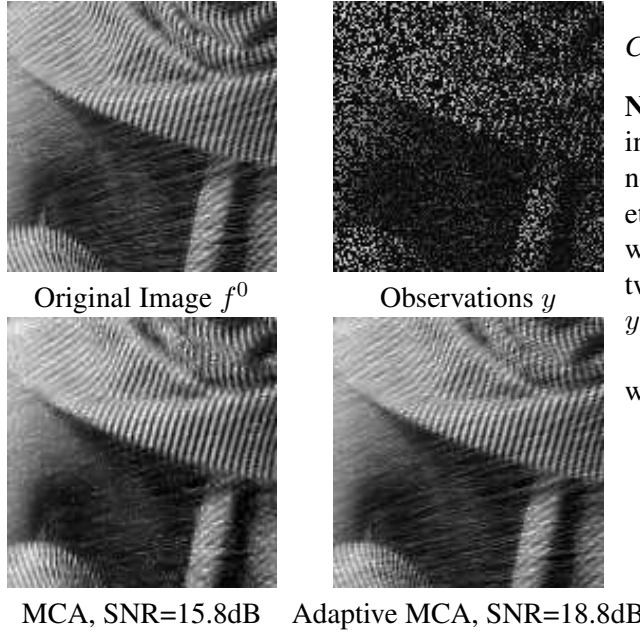

Original Image f^0 Observations y

MCA, SNR=15.8dB   Adaptive MCA, SNR=18.8dB

Fig. 16. *Comparison of inpainting using sparsity in a fixed dictionary of wavelets and local DCT (MCA), and using a wavelet basis and a learned dictionary (adaptive MCA).*

Application to computer graphics and vision. Texture synthesis is performed by computing a random synthesized image f whose patches are sparse in a given dictionary $\mathcal{D}$, see [74]. This dictionary $\mathcal{D}$ is learned from patches extracted from an exemplar f^0. The synthesis is obtained by initializing f to be a Gaussian white noise, and then iterating between the sparse patch coding (21) in $\mathcal{D}$ using $\pi_k = \pi_{x_k}(f)$ and averaging of the reconstructed patches

$$f(x) = \frac{1}{N_x} \sum_k \pi_k(x - x_k), \ \pi_k = \sum_{m=0}^{P-1} a_{k,m} \psi_m.$$

where N_x is the number of patches π_k that overlap at a point x_k.

Dictionary learning over sparse patches also allows to perform texture classification and segmentation [89], [74]. A dictionary is learned for each texture class, and a pixel is assigned to the class that performs the best sparse approximation around this pixel. A similar idea is applied to perform image recognition [58], where the dictionary is trained by enforcing both sparsity and discrimination of the classes. See [96] for a review of the application of dictionary learning to computer vision, including its application to face recognition.

C. Non-local Methods

Non-local Filtering. Non-local filtering have been introduced by Buades et al. [8] as an efficient denoising method. A spatially varying filter parameterized by λ computes $G_\lambda f(x) = \sum_y \lambda_{x,y} f(y)$ where $\lambda_{x,y} > 0$ characterizes the interaction between the features around point x and around point y.

An adaptive non-local filtering computes the weights $\lambda = \lambda(f)$ from the image f to process

$$\lambda_{x,y}(f) = \frac{\tilde{\lambda}_{x,y}}{Z_{\tilde{\lambda}}(x)} \quad \text{where} \quad (24)$$

$$\begin{cases} Z_{\tilde{\lambda}}(x) = \sum_y \tilde{\lambda}_{x,y}, \\ \tilde{\lambda}_{x,y} = \exp\left(-\frac{\|\pi_x(f) - \pi_y(f)\|^\alpha}{s}\right) \end{cases}$$

where $\alpha > 0$ and $s > 0$ controls the shape of the weights.

As s tends to zero, the filtering becomes related to texture synthesis method [37], [95], see Figure 7, third image from the left. Indeed, given an input image f, and starting from a random noise image $f^{(0)}$, applying the recursive scheme $f^{(k+1)} = G_{\lambda(f)} f^{(k)}$ for a small s produces a random image visually similar to f.

Non-local Spectral bases. This adaptive filtering is connected to adapted decomposition by introducing an adaptive non-local spectral basis [91], [72]. It is defined as the orthogonal eigenvectors basis $\mathcal{B}^\lambda = \{\psi_m^\lambda\}_{m=0}^{N-1}$ of the filtering operator $G_\lambda^* G_\lambda$

$$\psi_m^\lambda = \mu_m G_\lambda^* G_\lambda \psi_m^\lambda$$

and where $\mu_m > 0$ is the corresponding eigenvalue, assumed to be sorted in increasing order. The eigenvectors are a generalization of Fourier atoms to a spatially adaptive setting. Figure 17 shows examples of non-local spectral atoms adapted to the geometry of a cartoon image. They are oscillating near the boundary of the shape, to allow a better reconstruction of this singularity.
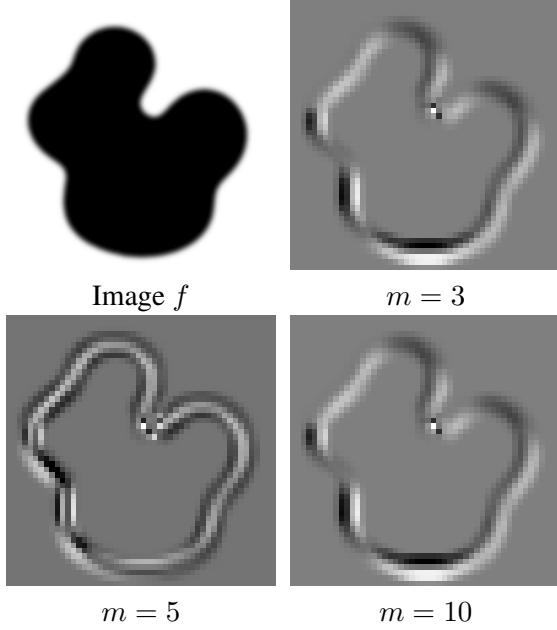
Fig. 17. *Example of non-local spectral vectors ψ_m^λ for several value of m.*

An adaptive denoising of a noisy observations $y = f^0 + w$ is obtained by computing the non-local graph $\lambda(y)$ and then thresholding the coefficients in this basis $H_T(y, \mathcal{B}^{\lambda(y)})$, see [72].

Non-local Regularization. The non-local filtering $G_\lambda f$ is related to the minimization of a non-local energy $\sum_{x,y} \lambda_{x,y} |f(x) - f(y)|^\alpha$, see for instance [50], [41], [21]. To tackle the regularization of an inverse problem $y = \Phi f^0 + w$, this energy is replaced by

$$J^\lambda(f) = \sum_{x,y} \lambda_{x,y} \|\pi_x(f) - \pi_y(f)\|^\alpha, \qquad (25)$$

see [43], [77]. For $\alpha = 2$, it is a generalization of the Sobolev prior to the non-local setting. For $\alpha = 1$, it extends the total variation.

For some inverse problems $y = \Phi f^0 + w$ such as deblurring [68], [53] or inpainting small holes [9], it is possible to estimate the graph from the observations $\lambda = \lambda(y)$. For generic inverse problems, the graphs λ needs to be optimized in parallel to the reconstruction. In this setting, the weights λ are constrained to be a probability distribution, so that $\sum_y \lambda_{x,y} = 1$ for each x. Following [43], the spread of the weights is controlled using a negative entropy $\mathcal{E}(\lambda) = s \sum_{x,y} w_{x,y} \log(w_{x,y})$ with weight $s > 0$.

Recovery from the noisy measurements $y = \Phi f^0 + w$ is obtained by minimizing

$$\min_{f \in \mathbb{R}^N, \lambda} E(f, \lambda) = \frac{1}{2\mu} \|y - \Phi f\|^2 + J^\lambda(f) + \mathcal{E}(\lambda),$$
$$(26)$$

as introduced in [43] for inpainting and in [77] for generic inverse problems. Note that this fits into the general adaptive framework (2).

The minimization of the non-convex problem (26) is performed by iteratively minimizing with respect to λ and to f. For a fixed f, the optimal weights minimizing (26) are defined as in (24). For a fixed λ, the optimal image is computed by minimizing a convex problem. One can shows that for $\alpha > 1$, this scheme converges to a stationary point of E defined in (26), see [77].

Figure (18) shows a comparison between this approach and total variation and wavelet sparsity for the recovery from missing random pixels.

CONCLUSION

This paper has reviewed several adaptive image representations, with an emphasis toward sparsity and non-local processing. The adaptivity to the geometry of edges and textures allows one to better represent these features. This geometry can be estimated directly from a clean or a noisy image to perform compression or denoising. Iterative algorithms are able to tackle more difficult inverse problems through the minimization of a non-convex energy. The resulting iterative estimation process progressively recovers both the geometric parameters and the image content.

REFERENCES

[1] P. Agarwal and S. Suri. Surface Approximation and Geometric Partitions. *SIAM J. Comput.*, 27(4):1016–1035, 1998.

[2] M. Aharon, M. Elad, and A.M. Bruckstein. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Proc.*, 54(11):4311–4322, 2006.

[3] J.-F. Aujol and A. Chambolle. Dual norms and image decomposition models. *International Journal of Computer Vision*, 63(1):85–104, juin 2005.

[4] I. Babuška and A. K. Aziz. On the angle condition in the finite element method. *SIAM Journal on Numerical Analysis*, 13(2):214–226, April 1976.

[5] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proc. of Siggraph 2000*, pages 417–424, 2000.

[6] S. Bougleux, G. Peyré, and L. Cohen. Image compression with geodesic anisotropic triangulations. *Proc. of ICCV'09*, pages 2343–2348, 2009.

[7] L. Breiman, J. H. Friedman, R. A. Olshen, , and C. J. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1984.
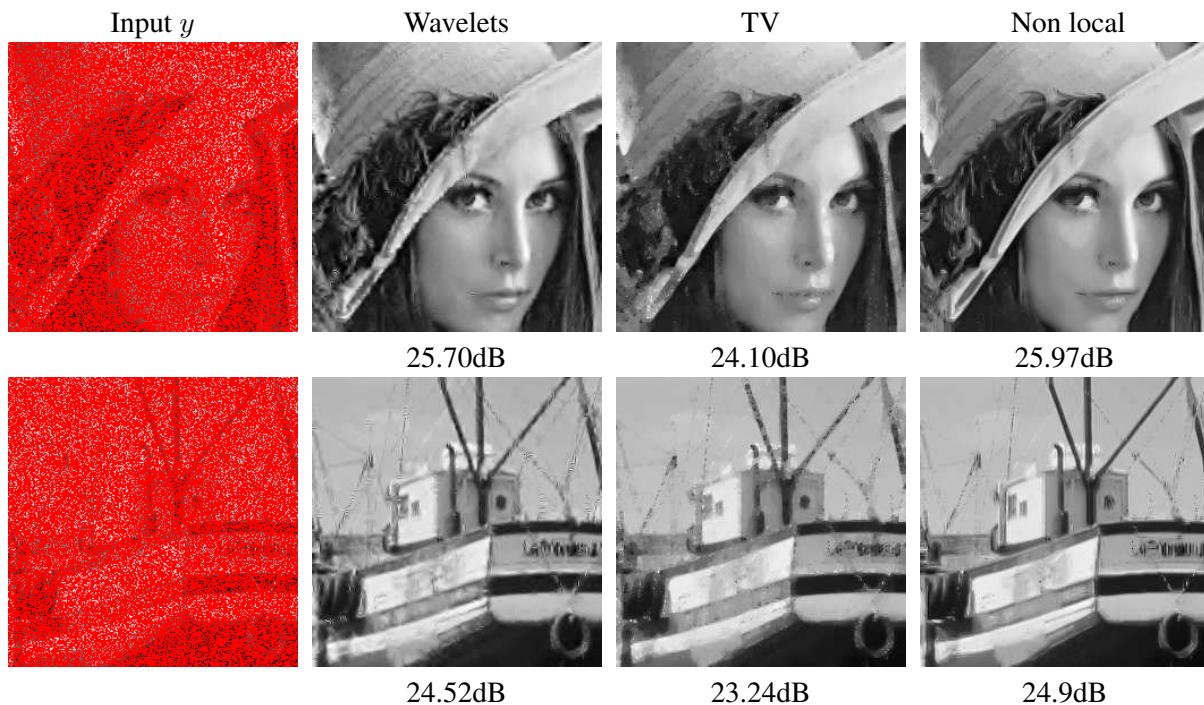
Input y	Wavelets	TV	Non local

	25.70dB	24.10dB	25.97dB

	24.52dB	23.24dB	24.9dB

Fig. 18. *Examples of inpainting with* 80% *of pixels.*

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

[9] A. Buades, B. Coll, J-M. Morel, and C. Sbert. Self similarity driven demosaicking. *IEEE Trans. Image Proc.*, 18(6):1192–1202, 2009.

[10] E. Candès and D. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Commun. on Pure and Appl. Math.*, 57(2):219–266, 2004.

[11] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Info. Theory*, 52(2):489–509, 2006.

[12] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying. Fast discrete curvelet transforms. *SIAM Multiscale Modeling and Simulation*, 3(5):861–899, 2003.

[13] E. J. Candès and D. L. Donoho. Recovering edges in ill-posed inverse problems: Optimality of curvelet frames. *Annals of Stat.*, 30(3):784–842, 2000.

[14] V. Chappelier and C. Guillemot. Oriented wavelet transform for image compression and denoising. *IEEE Trans. Image Proc.*, 15(10):2892–2903, oct 2006.

[15] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

[16] L. P. Chew. Guaranteed-quality mesh generation for curved surfaces. In *SCG '93: Proceedings of the ninth annual symposium on Computational geometry*, pages 274–280, New York, NY, USA, 1993. ACM.

[17] R. L. Claypoole, G. M. Davis, W. Sweldens, and R. G. Baraniuk. Nonlinear wavelet transforms for image coding via lifting. *IEEE Trans. Image Proc.*, 12(12):1449–1459, December 2003.

[18] A. Cohen and N. Dyn. Nonstationary subdivision schemes and multiresolution analysis. *SIAM J. Math. Anal.*, 27(6):1745–1769, 1996.

[19] A. Cohen, N. Dyn, F. Hecht, and J.-M. Mirebeau. Adaptive multiresolution analysis based on anisotropic triangulations. *to appear in Math. Comput.*, 2010.

[20] R. Coifman and V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Info. Theory*, IT–38(2):713–718, Mar. 1992.

[21] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102:7426–7431, May 2005.

[22] R. R. Coifman, G. Matviyenko, and Y. Meyer. Modulated Malvar-Wilson bases. *J. of Appl. and Comput. Harmonic Analysis*, 4(1):58–61, 1997.

[23] R.R. Coifman and D.L. Donoho. Translation-invariant de-noising. In *Wavelets and Statistics, Lecture Notes in Statistics*, pages 125–150. Springer Verlag, 1995.

[24] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Multiscale Modeling and Simulation*, 4(4):1168–1200, 2005.

[25] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. on Pure and Appl. Math.*, 57:1413–1541, 2004.

[26] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.*, 4(3):245–267, 1998.

[27] G. Davis, S. Mallat, and M. Avellaneda. Greedy adaptive approximation. *Journal of Constructive Approximation*, 13(1):57–98, 1997.

[28] L. Demanet and L. Ying. Wave atoms and sparsity of oscillatory patterns. *J. of Appl. and Comput. Harmonic Analysis*, 3(23):368–387, 2007.

[29] L. Demaret, N. Dyn, and A. Iske. Image compression by linear splines over adaptive triangulations. *Signal Processing Journal*, 86(7):1604–1616, 2006.

[30] L. Demaret, F. Friedrich, H. Fhr, and T. Szygowski. Multiscale wedgelet denoising algorithms. *Proceedings of SPIE, Wavelets XI*, 5914:X1–12, 2006.

[31] R. Distasi, M. Nappi, and S. Vitulano. Image compression by B-tree triangular coding. *IEEE Trans. Commun.*, 45(9):1095–1100, Sep 1997.

[32] D. Donoho. Wedgelets: Nearly minimax estimation of edges. *Annals of Stat.*, 27(3):859–897, 1999.

[33] D. Donoho. Compressed sensing. *IEEE Trans. Info. Theory*, 52(4):1289–1306, 2006.

[34] D. L. Donoho. Cart and best-ortho-basis: A connection. *Annals of Stat.*, 25(5):1870–1911, 1997.

[35] D. L. Donoho and I. Johnstone. Minimax estimation by wavelet shrinkage. *Annals of Stat.*, 26:879–921, 1998.

[36] N. Dyn, D. Levin, and S. Rippa. Data dependent triangulations for piecewise linear interpolation. *IMA J. Numer. Anal.*, 10(1):137–154, Jan. 1990.

[37] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, page 1033. IEEE Computer Society, 1999.

[38] A.A. Efros and W.T. Freeman. Image quilting for texture synthesis and transfer. *Proc. Siggraph '01*, pages 341–346, August 2001.

[39] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Proc.*, 15(12):3736–3745, 2006.

[40] M. Elad, M.A.T. Figueiredo, and Yi Ma. On the role of sparse and redundant representations in image processing. *Proceedings of the IEEE*, 98(6):972–982, 2010.

[41] A. Elmoataz, O. Lezoray, and S. Bougleux. Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing. *IEEE Trans. Image Proc.*, 17(7):1047–1060, 2008.

[42] K. Engan, S. O. Aase, and J. Hakon Husoy. Method of optimal directions for frame design. In *Proc. ICASSP '99*, pages 2443–2446, Washington, DC, USA, 1999. IEEE Computer Society.

[43] G. Facciolo, P. Arias, V. Caselles, and G. Sapiro. Exemplar-based interpolation of sparsely sampled images. *IMA Preprint Series # 2264*, 2009.

[44] J. C. Feauveau. Analyze multirésolution avec un facteur de rsolution $\sqrt{2}$. *J. Traitement du Signal*, 7(2):117–128, 1990.

[45] R. M. Figueras i Ventura, P. Vandergheynst, and P. Frossard. Low-rate and flexible image coding with redundant representations. *IEEE Trans. Image Proc.*, 15(3):726–739, March 2006.

[46] F. Friedrich, L. Demaret, H. Fuhr, and K. Wicker. Efficient moment computation over polygonal domains with an application to rapid wedgelet approximation. *SIAM J. Scientific Computing*, 29(2):842–863, 2007.

[47] H. Fuhr, L. Demaret, and F. Friedrich. Beyond wavelets: New image representation paradigms. *Survey article, in Document and Image Compression, M.Barni and F.Bartolini (eds)*, 2006.

[48] P.L. George, H. Borouchaki, P.J. Frey, P. Laug, and E. Saltel. Mesh generation and mesh adaptivity: theory, techniques. *Encyclopedia of computational mechanics,* *E. Stein, R. de Borst and T.J.R. Hughes ed., John Wiley & Sons Ltd.*, 2004.

[49] O. N. Gerek and A. E. Cetin. Adaptive polyphase sub-band decomposition structures for image compression. *IEEE Trans. Image Proc.*, 9(10):1649–1660, October 2000.

[50] G. Gilboa, J. Darbon, S. Osher, and T.F. Chan. Nonlocal convex functionals for image regularization. *UCLA CAM Report 06-57*, 2006.

[51] R. Gribonval and E. Bacry. Harmonic decomposition of audio signals with matching pursuit. *IEEE Trans. Signal Proc.*, 51(1):101–111, jan 2003.

[52] H.J. Heijmans, B. Pesquet-Popescu, and G. Piella. Building nonredundant adaptive wavelets by update lifting. *J. of Appl. and Comput. Harmonic Analysis*, 18(3):252–281, May 2005.

[53] S. Kindermann, S. Osher, and P. W. Jones. Deblurring and denoising of images by nonlocal functionals. *SIAM Mult. Model. and Simul.*, 4(4):1091–1115, 2005.

[54] H. Krim, D. Tuckier, S. Mallat, and D. Donoho. Near-optimal risk for best basis search. *IEEE Trans. Info. Theory*, 45(7):2225?–2238, 1999.

[55] E. Le Pennec and S. Mallat. Bandelet Image Approximation and Compression. *SIAM Multiscale Modeling and Simulation*, 4(3):992–1039, 2005.

[56] E. Le Pennec and S. Mallat. Sparse geometric image representations with bandelets. *IEEE Trans. Image Proc.*, 14(4):423–438, 2005.

[57] M. Lounsbery, T. D. DeRose, and J. Warren. Multiresolution analysis for surfaces of arbitrary topological type. *ACM Transactions on Graphics*, 16(1):34–73, 1997.

[58] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Supervised dictionary learning. *Proc. Advances Neural Information Processing Systems*, 2008.

[59] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Trans. Image Proc.*, 17(1):53–69, January 2008.

[60] S. Mallat. Geometrical grouplets. *J. of Appl. and Comput. Harmonic Analysis*, 26(2):161–180, 2009.

[61] S. Mallat. *A Wavelet Tour of Signal Processing, 3rd edition*. Elsevier, 2009.

[62] S. Mallat and G. Peyré. Orthogonal bandlets bases for geometric image approximation. *Commun. on Pure and Appl. Math.*, 61(9):1173–1212, 2008.

[63] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Proc.*, 41(12):3397–3415, 1993.

[64] S. Masnou. Disocclusion: a variational approach using level lines. *IEEE Trans. Image Proc.*, 11(2):68–76, February 2002.

[65] B. Matei and A. Cohen. *Nonlinear Subdivison Schemes : Applications to Image processing, in Tutorials on Multiresolution in Geometric Modelling*, pages 93–97. A. Iske, E. Quak and M.S. Floater ed., Springer Verlag, 2002.

[66] P. Maurel, J.F. Aujol, and G. Peyré. Locally parallel texture modeling. *to appear in SIAM Journal of Imaging Science*, 2010.

[67] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*. American Mathematical Society, Boston, MA, USA, 2001.

[68] M. Mignotte. A non-local regularization strategy for image deconvolution. *Pattern Recognition Letters*, 29(16):2206–2212, December 2008.

[69] J-M. Mirebeau and A. Cohen. Greedy bisection gener-

ates optimally adapted triangulations. Technical report, Laboratoire Jacques-Louis Lions, 2008.

[70] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive-field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, June 1996.

[71] N. Ouarti and G. Peyré. Best basis search in a non-stationary wavelet packets dictionary. *Proc. of ICIP 2009*, 2009.

[72] G. Peyré. Image processing with non-local spectral bases. *SIAM Multiscale Modeling and Simulation*, 7(2):703–730, 2008.

[73] G. Peyré. Manifold models for signals and images. *Computer Vision and Image Understanding*, 113(2):249–260, February 2009.

[74] G. Peyré. Sparse modeling of textures. *J. Math. Imaging Vis.*, 34(1):17–31, 2009.

[75] G. Peyré. Best basis compressed sensing. *IEEE Trans. Signal Proc.*, 28(5), 2010.

[76] G. Peyré. Texture processing with grouplets. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 4(32):733–746, 2010.

[77] G. Peyré, S. Bougleux, and L. D. Cohen. Non-local regularization of inverse problems. In *to appear in Inverse Problems and Imaging*, 2010.

[78] G. Peyré, J. Fadili, and J-L. Starck. Learning the morphological diversity. *to appear in SIAM Journal on Imaging Science*, 2010.

[79] G. Peyré and S. Mallat. Surface compression with geometric bandelets. *ACM Transactions on Graphics, (SIGGRAPH'05)*, 24(3):601–608, Aug. 2005.

[80] G. Peyré, E. Le Pennec, C. Dossal, and S. Mallat. Geometric estimation with orthogonal bandlet bases. In *SPIE Wavelet XII*, 2007.

[81] G. Plonka. The easy path wavelet transform: A new adaptive wavelet transform for sparse representation of two-dimensional data. *SIAM Multiscale Modeling and Simulation*, 7(3):1474–1496, 2009.

[82] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vision*, 40(1):49–70, 2000.

[83] J Portilla, V Strela, M Wainwright, and E P Simoncelli. Image denoising using a scale mixture of Gaussians in the wavelet domain. *IEEE Trans. Image Proc.*, 12(11):1338–1351, November 2003.

[84] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.

[85] J. Ruppert. A delaunay refinement algorithm for quality 2-dimensional mesh generation. *J. Algorithms*, 18(3):548–585, 1995.

[86] P. Schröder and W. Sweldens. Spherical Wavelets: Efficiently Representing Functions on the Sphere. In *Proc. of SIGGRAPH 95*, pages 161–172, 1995.

[87] A. Secker and D. Taubman. Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression. *IEEE Trans. Image Proc.*, 12(12):1530–1542, December 2003.

[88] R. Shukla, P.L. Dragotti, M. Do, and M. Vetterli. Rate distortion optimized tree structured compression algorithms for piecewise smooth images. *IEEE Trans. Image Proc.*, 14(3):343–359, 2005.

[89] K. Skretting and J.H. Husoy. Texture classification using sparse frame based representations. *EURASIP Journal on Applied Signal Processing*, 2006(1):102–102, 2006.

[90] W. Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *J. of Appl. and Comput. Harmonic Analysis*, 3(2):186–200, 1996.

[91] A. D. Szlam, M. Maggioni, and R. R. Coifman. Regularization on graphs with function-adapted diffusion processes. *Journal of Machine Learning Research*, 9:1711–1739, 2008.

[92] J. A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Trans. Info. Theory*, 52(3):1030–1051, 2006.

[93] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk. Wavelet-domain approximation and compression of piecewise smooth images. *IEEE Trans. Image Proc.*, 15(5):1071–1087, May 2006.

[94] D.M. Wang, L. Zhang, A. Vincent, and F. Speranza. Curved wavelet transform for image coding. *IEEE Trans. Image Proc.*, 15(8):2413–2421, August 2006.

[95] L-Y. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 479–488. ACM Press/Addison-Wesley Publishing Co., 2000.

[96] J. Wright, Yi Ma, J. Mairal, G. Sapiro, T.S. Huang, and Shuicheng Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, 2010.