Int J CARS (2016) 11:1765–1777
DOI 10.1007/s11548-016-1434-z

REVIEW ARTICLE

Patch-based models and algorithms for image processing: a review of the basic principles and methods, and their application in computed tomography

Davood Karimi[1] · **Rabab K. Ward**[1]

Received: 14 January 2016 / Accepted: 27 May 2016 / Published online: 10 June 2016

Abstract

Purpose Image models are central to all image processing tasks. The great advancements in digital image processing would not have been made possible without powerful models which, themselves, have evolved over time. In the past decade, "patch-based" models have emerged as one of the most effective models for natural images. Patch-based methods have outperformed other competing methods in many image processing tasks. These developments have come at a time when greater availability of powerful computational resources and growing concerns over the health risks of the ionizing radiation encourage research on image processing algorithms for computed tomography (CT). The goal of this paper is to explain the principles of patch-based methods and to review some of their recent applications in CT.

Methods We first review the central concepts in patch-based image processing and explain some of the state-of-the-art algorithms, with a focus on aspects that are more relevant to CT. Then, we review some of the recent application of patch-based methods in CT.

Results Patch-based methods have already transformed the field of image processing, leading to state-of-the-art results in many applications. More recently, several studies have proposed patch-based algorithms for various image processing tasks in CT, from denoising and restoration to iterative reconstruction. Although these studies have reported good results, the true potential of patch-based methods for CT has not been yet appreciated.

Conclusions Patch-based methods can play a central role in image reconstruction and processing for CT. They have the potential to lead to substantial improvements in the current state of the art.

Keywords Computed tomography · Low-dose CT · Reconstruction · Denoising · Restoration · Sparsity · Nonlocal means · Learned dictionaries

✉ Davood Karimi
karimi@ece.ubc.ca

Rabab K. Ward
rababw@ece.ubc.ca

1 2366 Main Mall, Vancouver, BC V6T 1Z4, Canada

Introduction

A central component in every computed tomography (CT) system is the image reconstruction algorithm. First CT scanners relied on iterative methods that aimed at recovering the unknown image as a solution of a system of linear equations. As the size of CT images grew, filtered-backprojection (FBP) methods became more common and they are still widely used. These methods require a large number of projections to produce a high-quality image, but they are faster than iterative methods. This speed advantage has become less significant as computers have become faster. On the other hand, there is a growing concern over the potential health risks of radiation exposure. As a result, there has been a renewal of interest in statistical and iterative image reconstruction methods because they can produce high-quality images under low-dose imaging conditions, albeit at a higher computational cost.

Another factor that has encouraged research on more sophisticated algorithms for CT is the introduction of new theories and methods in signal and image processing and applied mathematics. For instance, new optimization algorithms that have better convergence rates have led to fast iterative reconstruction algorithms. Another example is the developments in sparsity-based models. In these models, the image is transformed from its native representation in terms of pixel/voxel values into a different space where it has a more concise representation. Even though this is an old idea, recent years have witnessed the emergence of new theories and algorithms that have called for a reassessment of the potential of sparsity-based methods in CT. The use of edge-preserving regularization terms such as total variation (TV) for iterative CT reconstruction has received especial attention in recent years.

The focus of this paper is on *patch-based* models and algorithms. In patch-based methods, the units of operation are small image patches, which in the case of 3D images are also referred to as blocks. In the great majority of applications square patches or cubic blocks are used, even though other patch shapes can also be employed. For simplicity of presentation, we will use the term "patch" even when we are talking about 3D images. The number of pixels in a patch is usually on the order of tens or a few hundreds. A typical patch size is 8×8 pixels for 2D images or $8 \times 8 \times 8$ voxels for 3D images. In patch-based methods, the image is divided into small patches and each patch is processed individually. The final output image is formed by reassembling the individually processed patches. There are many reasons for focusing on small patches rather than on the whole image. First, because of the curse of dimensionality, it is much easier and more reliable to learn a model for small image patches than for very large patches or the whole image. Second, for many models, computations are significantly reduced if they are applied on small patches. Furthermore, research in the past decade has shown that working with small patches can result in highly efficient algorithms that outperform competing methods in a wide range of image processing tasks. For example, patch-based denoising methods are currently considered to be the state of the art, achieving close-to-optimal results.

Patch-based methods have been among the most heavily researched methods in image processing in recent years, and they have produced state-of-the-art results in many applications. Most of the published papers have focused on natural images. Comparatively, patch-based methods have received much less attention in CT. Given the success of these methods in various image processing applications, we believe that they have the potential to substantially improve the current state-of-the-art algorithms in CT. Therefore, the goals of this paper are twofold: first, to explain the central concepts in patch-based image processing and to briefly review some of the most important algorithms; second, to review some of the published work on patch-based image processing in CT. The next two sections will address the above two goals, in that order. A final section will present our concluding remarks.

Review of patch-based image processing methods

The word "patch-based" may be vague because it can refer to any algorithm that works with small image patches. For instance, image compression algorithms such as JPEG work on small patches. However, this word has recently been used to refer to certain classes of methods. In order to explain these methods, we first describe the two main frameworks in patch-based image processing. Most of the patch-based methods have their roots in one or both of these two frameworks.

Nonlocal patch-based image processing

Natural images contain abundant self-similarities. In terms of patches, this means that for every patch in a natural image, we can probably find many similar patches in the same image. Nonlocal patch-based methods exploit this self-similarity by finding/collecting similar patches and processing them jointly. Let us denote the noisy image by $x = x_0 + w$, where as before x_0 denotes the true image. We also denote the ith pixel of x by $x(i)$ and a patch centered on $x(i)$ by $x[i]$. We will use these notations in the rest of this paper. The nonlocal means (NLM) denoising algorithm estimates the value of the ith pixel in x_0 as a weighted average of the center pixels of all patches:

$$\widehat{x_0}(i) = \sum_{j=1}^{N} \frac{G_a\,(x[j] - x[i])}{\sum_{j=1}^{N} G_a\,(x[j] - x[i])} x(j) \tag{1}$$

where G_a denotes a Gaussian kernel with bandwidth a and N is the total number of patches [8]. The intuition behind NLM is that similar patches are likely to have similar pixels at their centers. In practical implementations of NLM, several patches that are sufficiently similar to $x[i]$ are found and used in computing $\widehat{x_0}(i)$.

The simple idea behind the NLM has proved to be an extremely powerful model for natural images. For denoising, NLM-based methods represent the state of the art. Extensions of the basic NLM include Bayeian/probabilistic extensions [52], spatially adaptive selection of the algorithm parameters [28], and the use of nonsquare patches to improve the denoising around the edges [23].

The idea of exploiting nonlocal patch similarities has been used for many other image processing tasks. For instance, nonlocal patch similarities have be used to develop highly effective regularizations for inverse problems and iterative image reconstruction algorithms [59,69,91]. Similar to NLM denoising, these algorithms exploit nonlocal patch similarities by introducing regularization terms that encourage similar patches to have similar pixels at their centers. Consider the inverse problem of estimating the unknown image x from measurements $y = Ax + w$, where w is the additive noise and matrix A represents the forward model (e.g., the projection matrix in CT). Many of these algorithms estimate x by solving an optimization problem that has the following general form:

$$\hat{x} = \underset{x}{\mathbf{argmin}} \, \|y - Ax\|_2^2 + \lambda \sum_i \sum_{j \in \mathcal{N}_i} s_{i,j} \|x(i) - x(j)\|_p \quad (2)$$

where $s_{i,j}$ is a measure of similarity between patches centered on the ith and jth pixels.

Because of space limitations, in this section we focused on image denoising and inverse problems, which are the most relevant to CT. However, the idea of exploiting nonlocal patch similarities has been applied to many image processing tasks, for example image enhancements [9], deblurring [50], and scale-up [71].

Patch-based image processing in learned overcomplete dictionaries

Learning overcomplete dictionaries

A signal $x \in \mathbb{R}^m$ is said to have a sparse representation in a dictionary $D \in \mathbb{R}^{m \times n}$ if it can be accurately approximated by a linear combination of a small number of its columns. Mathematically, $x \cong D\gamma$ such that $\|\gamma\|_0 \ll n$. Here, $\|\gamma\|_0$ denotes the number of nonzero entries of γ and is referred to as the ℓ_0-norm of γ. This means that only a small number of columns of D are sufficient for accurate representation of x. Columns of the dictionary D are commonly referred to as atoms. The ability to represent a signal as a linear combination of a small number of building blocks is a very powerful concept and is at the center of many algorithms in signal and image processing.

Traditionally, orthonormal bases such as wavelets have been used in image processing because they are easy to analyze and allow very efficient implementation. Over the past two decades, and especially in the past decade, there has been a significant shift of interest toward dictionaries that are adapted to a given class of signals via a learning strategy. The dictionaries obtained in this way lack the analytical and computational advantages of orthonormal bases, but they have much higher representational power. Therefore, they usually lead to superior results in many image processing tasks.

Suppose that we are given a set of training signals and would like to learn a dictionary for sparse representation of these signals. In image processing applications, each training signal is a vectorized patch. We stack these signals as columns of a matrix, which we denote with X. A dictionary can then be learned by solving the following optimization problem:

$$\underset{D \in \mathcal{D}, \Gamma}{\mathbf{minimize}} \, \|X - D\Gamma\|_F^2 + \lambda \|\Gamma\|_1 \quad (3)$$

In the above equation, and in the rest of this paper, Γ is the matrix of representation coefficients of the training signals in D, and $\mathcal{D}$ is the set of matrices whose columns have a unit Euclidean norm. The notations $\|\cdot\|_F$ and $\|\cdot\|_1$ denote, respectively, the Frobenius norm and the ℓ_1 norm. The constraint $D \in \mathcal{D}$ is necessary to avoid scale ambiguity. The first term in the objective function requires that the training signals be accurately represented by the columns of D and the second terms promotes sparsity, encouraging that a small number of columns of D are used in the representation of each training signal.

Equation (3) is only one possible formulation of the dictionary learning problem, and many variations of this formulation have been proposed (see e.g., [4,65]). The objective function in (3) is not jointly convex with respect to D and Γ, but it is convex with respect to D or Γ separately. Therefore, many dictionary learning algorithms adopt an alternating minimization approach. Examples of such algorithms are the method of optimal directions (MOD) [32] and the K-SVD algorithm [2]. There are also stochastic algorithms that are particularly designed for learning dictionaries from very large datasets [66]. Another class of dictionary learning algorithms include maximum-likelihood algorithms, which are in fact among the first proposed methods [54]. To encourage sparsity of the representation coefficients, these methods assume a sparsity-promoting prior such as a Cauchy or Laplace distribution for them. In recent years, fully Bayesian dictionary learning algorithms have also been proposed, which are highly flexible but computationally demanding [93].

There are many important variations of dictionary learning. Due to space limitations, here we briefly mention three of them. The first is the structured dictionary learning, where the interaction between the dictionary atoms is also learned/imposed. Common structures are tree structures [39] and grid structures [46]. The second variation is multiscale dictionary learning, extending the basic dictionary learning scheme to consider different patch sizes [68]. The third important variation includes dictionaries that have a fast application. As we mentioned above, compared with analytical dictionaries such as wavelets, learned dictionaries are more costly to deploy. Therefore, there are dictionary structures that have been proposed to reduce the computational cost [1,73]. As a final remark, we should mention that preprocessing of the training patches significantly influences the types of structures that emerge in the learned dictionary and the performance of the dictionary in practice. Commonly used preprocessing operations include centering, variance normalization, and whitening. The overall effect of these operations is to amplify the high-frequency structures such as edges, resulting in more high-frequency patterns in the learned dictionary.

Image processing using learned overcomplete dictionaries

Learned dictionaries have been employed in various image processing applications. Below, we describe the basic formulations for image denoising and scale-up. Not only these are among the most successful applications of learned dictionaries, they are also very instructive in how these dictionaries are used to accomplish various image processing tasks.
Image denoising

Suppose that we have measured a noisy image $x = x_0 + w$, where x_0 is the true underlying image and w is the white Gaussian noise. The prior assumption in denoising using a dictionary D is that every patch in the image has a sparse representation in D. Using this prior on every patch in the image, the maximum a posteriori (MAP) estimation of the true image can be found as the solution of the following problem [30]:

$$\left\{\widehat{x_0}, \{\widehat{\gamma_i}\}_{i=1}^N\right\} = \underset{z,\{\gamma_i\}_{i=1}^N}{\mathbf{argmin}}\, \lambda\|z - x\|_2^2 + \sum_{i=1}^{N}\left(\|R_i z - D\gamma_i\|_2^2 + \|\gamma_i\|_0\right) \tag{4}$$

where R_i represents a binary matrix that extracts and vectorizes the ith patch from the image. This is a common notation that is used to simplify the presentation of this type of equations, and we will use it in the rest of this paper. N is the total number of extracted patches. It is common to use overlapping patches to avoid discontinuity artifacts at the patch boundaries. The common approach to solving this optimization problem is an approximate block-coordinate minimization [31].

Image scale-up (super-resolution)

Suppose x_h is a high-resolution image. A blurred low-resolution version of this image can be modeled as $x_l = SHx_h + w$, where H and S are blur and downsampling operators. The goal is to recover the high-resolution image. This problem is usually called the image scale-up or super-resolution. The algorithm proposed in [88] learns two dictionaries (D_h, D_l) for representation of the patches of the high-resolution and low-resolution images, respectively. The basic assumption in this algorithm is that sparse representation of a low-resolution patch in D_l is identical to the sparse representation of its corresponding high-resolution patch in D_h. These dictionaries can be learned by solving:

$$\underset{D_h, D_l, \Gamma}{\mathbf{minimize}}\ \frac{1}{m_h}\|X^h - D_h\Gamma\|_F^2 + \frac{1}{m_l}\|X^l - D_l\Gamma\|_F^2 + \lambda\|\Gamma\|_1 \tag{5}$$

where X^h and X^l represent the matrices of training signals, and m_h and m_l are the length of the high-resolution and low-resolution training signals and are included in the objective function to balance the two terms. The above algorithm obtained surprisingly good results [88]. Later studies showed that even better results can be achieved by relaxing the relation between the representation coefficients of low-resolution and high-resolution patches [37,83].

Image scale-up is an example of what has been called "task-driven dictionary learning." In task-driven dictionary learning, the goal is to learn the dictionary not only for sparse representation of the signal, but also so that it can be employed for prediction of some target variable (i.e., classification or regression). For the image scale-up problem, for instance, the target variables are the pixel values of the high-resolution patch. A general formulation of task-driven dictionary learning and an efficient optimization problem were proposed in Mairal et al. [64]. Learned dictionaries have proved highly effective in many other applications as well. Examples include deblurring [19], classification [72], and restoration [41], to name only a few.

Other patch-based methods in image processing

Many of the more elaborate patch-based algorithms have combined the ideas of learned dictionaries and nonlocal similarities to enjoy the benefits of both methods. The first algorithm to explicitly follow this approach was "the nonlocal sparse model" proposed in [67]. This method collects similar patches, as in NLM. However, unlike NLM that performs a weighted averaging, the nonlocal sparse model uses sparse coding of similar patches in a learned dictionary. The basic assumption in this algorithm is that similar patches use similar dictionary atoms in their representation.

The above idea has been explored by many studies in the recent years [24,27]. Although the details of these algorithms are different, the main ideas can be simply explained in terms of the nonlocal patch similarities and sparse representation in learned dictionaries. A surprising similarity of many of these algorithms is the use of simple dictionaries (e.g., PCA dictionaries) for clustered patches. BM3D algorithm, which is arguably the best image denoising algorithm, uses orthonormal bases [20,63]. It collects similar patches and filters them jointly using orthogonal DCT dictionaries, although the details of the algorithm are more complex.

Current state-of-the-art image denoising methods are based on both concepts of learned dictionaries and nonlocal patch similarities. Although originally denoising methods based on sparse representation in learned dictionaries and those based on nonlocal patch similarities were developed independently, later studies obtained better results by combining the two approaches. However, studies have shown that the idea of exploiting nonlocal patch similarities is more powerful than the idea of sparse representation in learned dictionaries [12,52]. In general, these state-of-the-art methods achieve results that are very close to optimal on many natural images. However, on very difficult images, such as images with nonstationary and complex texture, their performance is much lower [11].

Patch-based image processing in the presence of Poisson noise

The noise in CT projection measurements can be approximated as a Poisson or Poisson-plus-Gaussian noise. Unfortunately, the great majority of the patch-based algorithms have been proposed for uniform Gaussian noise. Only very recently, a number of works have considered the case of Poisson noise. Therefore, we decided to review them here in a separate section.

An important first obstacle facing the application of patch-based methods to the case of Poisson noise is assessment of patch similarity. A reliable measure of patch similarity is needed for both nonlocal methods (to find similar patches) and dictionary-based methods (to compute the sparse representations). For very low-count Poisson measurements, one study suggested that the earth-mover's distance (EMD) is a good measure patch dissimilarity and that EMD can be estimated by passing the noisy patches through a Gaussian filter and then computing the Euclidean distance [34]. Another study compared several different patch similarity measures for Poisson noise and found that the generalized likelihood ratio (GLR) was the best [26]. GLR has many desirable theoretical properties that make it very appealing as a patch similarity measure [22]. For two Poisson-distributed random variables, x_1 and x_2, GLR is computed as:

$$\mathcal{L}_G\left(x_1, x_2\right)=\frac{\left(x_1+x_2\right)^{x_1+x_2}}{2^{x_1+x_2} x_1^{x_1} x_2^{x_2}} \tag{6}$$

For two image patches, assuming that the noise in pixels are independent, GLR is a simple extension of the above equation. In brief, GLR approximates the ratio of the probability that the two signals (e.g., two image patches) are noisy realizations of the same noise-free signal to the probability that the two signals are realizations of two different noise-free signals.

In Deledalle et al. [22], GLR was also compared with six other criteria for nonlocal patch-based denoising of images with Poisson noise. It was found that using GLR led to the best denoising result when the noise was strong. The algorithm used in [22] for nonlocal filtering was as follows:

$$\widehat{x_0}(i)=\sum_{j=1}^{N} \frac{\mathcal{L}_G\left(x[j]-x[i]\right)^{1/h}}{\sum_{j=1}^{N} \mathcal{L}_G\left(x[j]-x[i]\right)^{1/h}} x(j) \tag{7}$$

which includes the tuning parameter h.

A nonlocal patch-based denoising algorithm for Poisson noise was suggested in [25]. A main feature of this algorithm is that the patch similarity weights are computed from the original noisy image as well as from a prefiltered image. Another study used GLR to develop a k-medoids denoising algorithm [10] and showed that it outperformed the nonlocal Poisson denoising method of [22] in some tests. The k-medoids algorithm is a special case of the dictionary-based approach, the difference being that the k-medoids uses only one atom for representation of each patch. The reason why only one atom was used for representation of each patch was the difficulties in sparse coding under the Poisson noise. These difficulties have been discussed in [29,34], and greedy sparse coding algorithms have been proposed for solving this problem. The authors of [29] applied their algorithm for denoising of images with Poisson noise with a wavelet dictionary and achieved impressive results.

A true dictionary-based denoising algorithm for images with Poisson noise was suggested in [34]. In that study, a global dictionary is learned from a set of training data. Then, for a given noisy image to be denoised, the algorithm first clusters similar patches together. All patches in a cluster are denoised together via simultaneous sparse representation in D. This means that patches that are clustered together are forced to share similar dictionary atoms in their representation. Experiments showed that this method was comparable with or better than competing methods. A similar approach that also combines the ideas of learned dictionaries and nonlocal filtering is proposed in [75]. In this approach, k-means clustering is used to group similar image patches. A dictionary is learned for each cluster of similar patches using the Poisson-PCA algorithm. The algorithm showed good performance under low-count Poisson noise.

Review of published research on patch-based processing in CT

In this section, we review some of the recent applications of patch-based methods in CT by dividing them into three categories: (1) preprocessing methods, which work on the projection measurements, i.e., the sinogram, (2) iterative reconstruction methods, and (3) post-processing methods, whose goal is to restore or denoise the reconstructed image.

Preprocessing methods

Very few studies have applied patch-based methods on CT projections. This is partly because the great majority of the patch-based image processing algorithms have been proposed with the assumption of uniform Gaussian noise. Moreover, the algorithms that have been proposed for Poisson noise are very recent and have not been yet absorbed by researchers working on CT.

A patch-based sinogram denoising algorithm was proposed in [76]. A fixed DCT dictionary was used. However, the shrinkage rule used for denoising was learned from training data. The denoised projections were used to reconstruct the image with FBP. A patch-based processing using learned shrinkage functions was then applied on the reconstructed image. It is well known that image features are blurred in the projection measurements. Therefore, algorithms working with the projections usually have difficulty preserving image features. In order to tackle this challenge, in [76] the shrinkage rules for denoising the projection measurements were learned by minimizing a cost function based on the error in the *image* domain, not the projection domain. Furthermore, the cost function considered the error not only in the intensity value of the image, but also in the image gradient. The intuition behind this choice is that this formulation of cost function is very sensitive to small and low-contrast edges. Therefore, the shrinkage rules for denoising of the projections are learned such that image features are preserved. The results of the study showed that this rather simple algorithm outperformed some of the simple iterative CT reconstruction algorithms.

Recently, a novel dictionary-based sinogram denoising algorithm was proposed in [45]. This algorithm is based on simultaneous sparse representation of blocks extracted from stacked projections. In order to exploit the correlation between neighboring pixels in each projection as well as the correlation between pixels in neighboring projection views, 2D cone-beam projections are stacked together to form a 3D image. Then, overlapping blocks are extracted from this 3D image. These blocks are then clustered such that each cluster contains very similar blocks. The main assumption in the algorithm is that all blocks in a cluster must share the same dictionary atoms in their sparse representations. Experiments showed that this sinogram denoising algorithm achieved results that were better than or at least comparable with the results obtained by the state-of-the-art algorithms (such as BM4D). This algorithm was also shown to be very fast. The main reason behind the speed of this algorithm was that all overlapping blocks were clustered only once. Moreover, this algorithm proposed a novel strategy to project high-dimensional blocks onto a very low-dimensional space, where clustering could be performed at a much lower cost.

Use of learned dictionaries for interpolation (i.e., upsampling) of the CT projection measurements has been proposed in at least two studies [44,55]. The approach used by both of these studies was very similar to the general inpaining approach proposed in [30]. The work in [44] stacked 2D cone-beam projections to form a 3D image and worked with blocks extracted from this 3D image to exploit the correlation between neighboring detectors as well as the correlation between neighboring projections. The results of both of these studies showed a substantial improvement in the quality of the images reconstructed with FBP. The study in [55] showed that this method outperformed sinogram interpolation using splines.

A challenge for all sinogram denoising/restoration algorithms is preservation of fine image detail. To address this challenge, one study suggested learning a dictionary for sparse representation of sinogram patches by considering not only the sinogram-domain error but also the error in the image domain [77]. Specifically, first a dictionary (D_1) is learned considering only the error in the sinogram domain. Let us denote the CT image and its sinogram with x and y, respectively. Then D_1 is found by solving:

$$\left\{D_1, \hat{\Gamma}\right\} = \underset{D,\Gamma}{\textbf{argmin}} \|\Gamma\|_0 \quad \textbf{subject to}: \|D\Gamma_i - R_i y\|_2^2 \leq C\sigma_i \quad \forall i \tag{8}$$

This dictionary is then further optimized by minimizing the reconstruction error in the image domain:

$$D_2 = \underset{D}{\textbf{argmin}} \left\| \mathcal{FBP}\left(\sum_i \left(R_i^T R_i\right)^{-1} \sum_i R_i^T D\hat{\Gamma} \right) - x \right\|_{Q,2}^2 \tag{9}$$

where $\mathcal{FBP}$ denotes the CT reconstruction algorithm. Note that for finding D_2 we keep the sparse representations fixed and find a dictionary that leads to a better reconstruction of the image x. The notation $\|\cdot\|_{Q,2}$ denotes a weighted ℓ_2 norm, where the weights Q are chosen such that more weight is given to low-contrast features. Then, noisy projections of a new object/patient, y_{noisy}, are denoised in two steps. First, sparse representations of patches of y_{noisy} in D_1 are obtained. Denoting this with $\hat{\Gamma}$, the final denoised sinogram is obtained by solving the following problem, which uses D_2:

$$y_{\text{denoised}} = \underset{y}{\textbf{argmin}}\, \lambda \|y - y_{\text{noisy}}\|_W^2 + \sum_i \|D_2\hat{\Gamma}_i - R_i y\|_2^2 \tag{10}$$

where W are weights to account for the signal-dependent nature of the noise. Experiments with simulated and clinical data showed promising results [77].

Iterative reconstruction methods

In recent years, several iterative image reconstruction algorithms involving regularizations in terms of image patches have been proposed for CT. In general, these algorithms have reported very promising results. However, a convincing comparison with other classes of iterative reconstruction algorithms such as those based on edge-preserving regularizations is still lacking.

A dictionary-based iterative reconstruction algorithm was proposed in [84]. It estimates the image as a solution of the following optimization problem:

$$\underset{x,D,\Gamma}{\textbf{minimize}} \sum_i w_i([Ax]_i - y_i))^2 + \lambda \left(\sum_k ||R_k x - D\Gamma_k||_2^2 + \nu_k ||\Gamma_k||_0 \right) \tag{11}$$

where A is the projection matrix and w_is are noise-dependent weights. The first term in the objective function encourages measurement consistency. The remaining terms constitute the regularization, which are similar to the dictionary learning problem in (3). This problem is solved via alternating minimization. Minimization with respect to x is carried out using the separable paraboloid surrogate method. The problem with this approach, however, is that it requires access to the individual elements of the projection matrix, which can be a major limitation for large 3D CT. Minimization with respect to D and Γ is carried out similar to the K-SVD algorithms. Alternatively, the dictionary can be learned in advance. This will remove D from the list of the optimization variables in (11), substantially simplifying the problem. Experiments showed that both of these approaches lead to very good reconstructions, outperforming a TV-based algorithm [84].

Formulations similar to (11) were shown to be superior to TV-based reconstruction and other standard iterative reconstruction algorithms in electron tomography [3,58]. Another study first learned a dictionary from training images, but for image reconstruction step did not include the sparsity term in the objective function [33]. In other words, only the first two terms in the objective function in (11) were considered. That study found superior reconstructions with learned dictionaries compared with a DCT basis. Another study used an optimization formulation similar to (11), but used box-splines for image representation [74]. The results of that study showed that this dictionary-based algorithm achieved much better reconstructions than a wavelet-based algorithm.

There have also been some studies on dual-dictionary methods [60,92] that rely on two dictionaries. The atoms of the two dictionaries are in one-to-one correspondence. One of the dictionaries is composed of patches from CT images reconstructed from a small number of views, while the second dictionary contains the corresponding patches from a high-quality image. The strategy here is to find the sparse code of the image to be reconstructed in the dictionary of few-view patches and recover an estimate of the high-quality patch by multiplying this sparse code with the high-quality dictionary. This approach has been reported to achieve better results than TV-based reconstruction [60].

A different dictionary-based reconstruction algorithm was suggested in [79]. In this algorithm, first a dictionary D is learned by solving a problem of the following form:

$$\begin{aligned} &\underset{D,\Gamma}{\textbf{minimize}} \; \|X - D\Gamma\|_F^2 + \lambda \|\Gamma\|_1 \\ &\textbf{subject to}: \quad D \in \mathcal{D} \quad \& \quad H \in \mathbb{R}_+ \end{aligned} \tag{12}$$

where X is the matrix of training signals. In other words, each column of X is one vectorized image patch. For reconstruction, however, instead of using overlapping patches a novel regularization term is introduced to avoid the blocking artifacts at the patch borders. The regularization has the form $\|Lx\|_2^2$ where the matrix L computes the directional derivatives across the patch boundaries. Evaluations showed that this algorithm reconstructs high-quality images, preserving fine textures that are smeared by TV-based reconstruction. A follow-up paper studied the sensitivity of this algorithm to such factors as the scale and rotation of features in the training data [78].

An iterative reconstruction algorithm that combines sparse representation of image patches with sinogram smoothing was proposed in [80]:

$$\underset{x,y,\Gamma}{\textbf{minimize}} \; \|y - \bar{y}\| + \alpha\, y^T W y + \beta \|Ax - y\|_2^2 + \lambda \left(\sum_k ||R_k x - D\Gamma_k||_2^2 + \nu_k ||\Gamma_k||_0 \right) \tag{13}$$

The first two terms, where $\bar{y}$ is the measured noisy sinogram, represent the sinogram Markov random field model. That study also suggested interesting variations of the objective function in (13), but the experimental evaluations are limited.

Two interesting variations of dictionary-based iterative CT reconstruction are the tensor-based algorithm [81] and the method based on sparsifying transforms [70]. Tensor-based methods treat the image patches or blocks in their original form, i.e., without vectorizing them. Therefore, they are expected to better exploit the correlation between adjacent pixels. In Tan et al. [81], a tensor-based algorithm was compared with a standard dictionary for dynamic CT reconstruction and it was shown to be slightly better. Sparsifying transforms, on the other hand, are based on the analysis model for sparsity. In the analysis model, instead of the relation $x = D\gamma$ that we discussed in this paper, we have $Dx = \gamma$. In other words, D acts as an operator on the signal (e.g., the image patch) to find the representation coefficients γ. The study in [70] showed that iterative reconstruction based on sparsifying transforms leads to results that are comparable with TV-based reconstruction, while also being marginally faster.

In recent years, there has also been a growing attention to the potential of regularization in terms of nonlocal patch priors for iterative CT reconstruction. Most of these studies have suggested to recover the image as a solution of an optimization problem of the form:

$$\hat{x} = \underset{x}{\mathbf{argmin}} \, \|y - Ax\|_2^2 + \lambda J_{\mathrm{NL}}(x) \tag{14}$$

where $J_{\mathrm{NL}}(x)$ is the regularization in terms of patch similarities. Various forms have been suggested for $J_{\mathrm{NL}}(x)$, many of which are very similar to that in (2).

One study computed the patch similarity weights from a FBP-reconstructed image while accounting for the local noise variance and minimized (14) using gradient descent [59]. The recovered image had a better visual and objective quality than a TV-based algorithm. A simple block-coordinate minimization algorithm for solving this type of reconstruction problem was suggested in [38], where minimization of the measurement misfit term is carried our using projection onto convex sets. One study adopted a Bayesian formulation by modeling the measurement consistency as the log-likelihood of the Poisson-distributed photon counts [14]. Another study developed a NLM-type regularization for perfusion CT that relies on a high-quality prior image [62]. The patch similarity weights were computed between patches of the reconstructed image (from a low-dose scan) and those of the prior image. A similar approach was suggested in [89]. A more general algorithm that computes the patch similarity weights from the reconstructed image itself was proposed in [90]. Both of the above studies reported very good reconstruction results, outperforming more conventional regularizations such as Gaussian Markov random field.

Finally in this section, we briefly review some of the representative work on patch-based iterative reconstruction algorithms for dynamic CT. In dynamic CT, several successive images of the same patient are reconstructed. Therefore, there is abundant temporal correlation (i.e., correlation between successive images) in addition to the spatial correlation within each of the images in the sequence. A reconstruction algorithm with nonlocal patch-based regularization was proposed in [47]. The proposed regularizer for the kth frame of the image is as follows:

$$\begin{aligned} J(x_k) &= \sum_i \sum_{j \in \mathcal{N}_i} G_a(x_k[i] - x_k[j]) |x_k(i) - x_k(j)|^2 \\ &\quad + \sum_i \sum_{l \in \{1,2,\ldots,K\} \setminus k} \sum_{j \in \Delta_i} \\ &\quad \times G_a(x_l[i] - x_k[j]) |x_l(i) - x_k(j)|^2 \end{aligned} \tag{15}$$

where, similar to Eq. (1), G_a denotes a Gaussian kernel with bandwidth a. The first term is a spatial regularization term, where $\mathcal{N}_i$ is a neighborhood around the ith pixel. The second term, where K is the total number of frames, is a temporal regularization that involves patches from all other frames in the image sequence. In this term, Δ_i is a neighborhood whose spatial size if prefixed but whose temporal extension is found for each pixel such that the probability of finding patches with similar structure is increased. A similar algorithm was proposed in [48] for the case when a high-quality prior image is available. The results of experiments with simulated and real data show that this algorithm achieves very good reconstructions.

A temporal nonlocal means (TNLM) was proposed in [40, 82] using a regularization of this form:

$$J_{\mathrm{NL}}(\{x_k\}) = \sum_{k=1}^{K} \sum_i \sum_j s_{i,j} \left(x_k(i) - x_{k+1}(j)\right)^2 \tag{16}$$

Only inter-image patch similarities are taken into account and not the intra-image patch similarities. The reasoning is that using patches from the same image will amplify the streak artifacts, while using patches from neighboring images will suppress the artifacts. In Jia et al. [40] a similar post-processing algorithm was proposed, where each image in the sequence is first reconstructed from its corresponding projections and then it is post-processed using an algorithm that includes the regularization function in (16).

Post-processing methods

Many of the patch-based algorithms that have been proposed for CT fall into the category of post-processing methods. This is partly because most of the patch-based algorithms that have been developed for CT are directly based on methods that have been proposed for natural images. Because general patch-based image processing algorithms mostly include denoising and restoration algorithms, they are more easily extended as post-processing methods for CT. Moreover, patch-based methods are computationally intensive. Therefore, it is easier to deploy them as one-shot post-processing algorithms than as iterative reconstruction algorithms.

A large number of dictionary-based algorithms have been proposed for CT denoising. The basic denoising algorithm described in "Image processing using learned overcomplete dictionaries" section was used for denoising of abdomen

[18], head [7,16], and micro-CT images [42] with promising results. This simple algorithm resulted in effective suppression of noise and artifacts and a marked improvement in the visual and objective image quality. Nonlocal means methods have also been applied for CT image denoising. An early example is [49], where the effect of patch size, smoothing strength, and size of the search window was investigated. Among the findings of that study with lung and abdomen CT images was that one can choose very small search windows and still achieve effective denoising. Moreover, with a small search window, the patch size also has to be small to ensure effective denoising around the edges. Another study found that with a basic NLM denoising, the tube current setting can be reduced to 1/5 of that in routine abdominal imaging without jeopardizing the image quality [13].

An NLM-type algorithm specially tailored to image-guided radiotherapy was proposed in [87]. Since in this scenario a patient is scanned multiple times, the first scan is performed with standard dose and later scans with much reduced dose. The low-dose images are denoised by finding similar patches in the standard-dose image. The NLM-type algorithm suggested for CT perfusion imaging and angiography in [61] registers the standard-dose image to the low-dose image. The low-dose image is then denoised using patches extracted from the standard-dose image. One study proposed a method for approximating the local noise level and adapted the strength of the NLM denoising accordingly [57]. Their results show that this algorithm effectively suppresses the noise without degrading the spatial resolution. Using speed-up techniques such as those in [21] and implementation on GPU, they are able to process large 3D images in a few minutes. Another study estimated the local noise variance by reconstructing separate images from even- and odd-numbered projections and analyzing the difference of the images [6].

Recently, several patch-based algorithms have been proposed specifically for suppressing artifacts in CT images. One algorithm decomposes the artifact-full image into high-frequency bands in the horizontal, vertical, and diagonal directions and computes the sparse representation of patches of these bands in three "discriminative dictionaries," that include atoms learned to represent artifacts and genuine image features [15]. Artifacts are suppressed by setting to zero the large coefficients that correspond to the artifact atoms. The results of this study on artifact-full CT images are impressive. A very common type of artifact that arises in CT images reconstructed from a small number of projections is streak artifact. A dictionary-based algorithm for suppressing this type of artifact was proposed in [43]. This algorithm is based on building two dictionaries, one for artifact-full images and another for clean artifact-free images. Sparse representations of artifact-full and clean image blocks in their corresponding dictionaries are assumed to have a linear relation, and this relation is learned from training data using a formulation very similar to that described for image scale-up in "Image processing using learned overcomplete dictionaries" section. The results of this study were also very promising. A nonlocal patch-based method for suppressing streak artifacts that arise when the number of projections is small was proposed in [86]. It registers a high-quality reference image to the few-view image that is marred by streak artifacts. The registered image is then used to simulate a few-view image. Suppression of streak artifacts is achieved by matching patches between the original artifact-full image and the simulated few-view image, and then using the corresponding high-quality patches from the reference image. This algorithm is extended in [85] to be used when a prior scan from the same patient is not available but a rich database of images is available. Both methods substantially reduced the streak artifacts in images reconstructed from less than 100 projections. Another study used a wavelet dictionary for artifact suppression, which was then followed by NLM denoising [17].

A major challenge facing the application of patch-based algorithms for large CT images is the issue of the computational time. Although we discuss this challenge here under the post-processing methods, it applies equally to pre-processing methods and, especially, iterative reconstruction algorithms. Obvious approaches to reducing the computational load include using faster hardware such as GPU (which is becoming a routine practice [6,57]) and working with small 2D patches (which is very suboptimal for 3D images [56,73]). However, since the focus of this review is on algorithmic aspects, below we briefly review some of algorithmic speedup strategies.

For dictionary-based methods, the most computationally intensive step is the sparse coding of image patches. To speed-up this operation, several "structured dictionaries" have been proposed. For example, in the "double sparsity" model the learned dictionary has the form $D = \Phi A$, where Φ is an orthonormal basis and A is a sparse matrix [73]. This dictionary reduced the denoising time for 3D CT images by factors of around 30, while also improving the denoising performance. A similar idea is the separable dictionary where D is the Kronecker product of two smaller dictionaries [36]. This dictionary reduces the complexity of sparse coding from $\mathcal{O}(n)$ to $\mathcal{O}(\sqrt{n})$. Clustered/multilevel dictionaries have also been proposed for fast processing of 3D CT images [42,56]. There are also fast sparse coding algorithms, some of which have been successfully applied to CT. A discussion of these algorithms is beyond the scope of this paper, but two notable mentions include [35,53]. For nonlocal patch-based algorithms, the computational bottleneck is the search for similar patches. There has been great progress in reducing the cost of this step. Successful algorithms include randomized methods [5] and tree-based methods [51].

Conclusions

Patch-based models have emerged as one of the most successful models for natural images. Their practical utility has been demonstrated by hundreds of studies in recent years. In this paper, we reviewed some of the main concepts and methods in this field. Use of learned overcomplete dictionaries for sparse representation of image patches and use of nonlocal patch similarities are at the core of much of the ongoing research in image processing. On many image processing tasks, these seemingly simple ideas have led to algorithms that are superior or at least on a par with other competing methods.

We also reviewed some of the published research on the application of patch-based methods for reconstruction and processing of CT images. Most of these studies have obtained very good results. However, the amount of research on the application of these methods in CT has been far below the expectation. We think that any observant reader who is familiar with the challenges of reconstruction and processing of CT images will acknowledge that there is an immense potential for patch-based methods to improve the current state-of-the-art algorithms in CT.

In terms of the preprocessing algorithms, there has been only a couple of published papers on patch-based algorithms. This is partly because most of the patch-based models and algorithms have been originally proposed for uniform Gaussian noise. Only very recently, similar methods for the case of Poisson noise have started to appear. Nonetheless, even with the current tools, powerful patch-based preprocessing algorithms can be developed for CT. Some of the patch-based methods that we reviewed in "Patch-based image processing in the presence of Poisson noise" section were applied on very low-count Poisson noise and achieved impressive results. This can be extremely valuable for low-dose CT that is of especial interest in clinical settings.

Patch-based iterative CT reconstruction algorithms have reported very promising results. Some of the most promising results have been reported for dynamic CT, where there is abundant temporal as well as spatial patch similarities. Nonetheless, many of the proposed algorithms have been applied on 2D images and in some cases it is not clear how they can be applied to large 3D reconstruction. Moreover, little is known about the robustness of these algorithms in terms of the trained dictionary. To the best of our knowledge, there has been only one such study that investigated the effect of the scale and orientation of features in the training images on the performance of the dictionary for iterative reconstruction [78]. Finally, a comprehensive comparison of patch-based algorithms with new iterative reconstruction algorithms based on edge-preserving regularizations is still lacking.

Patch-based post-processing of CT images for denoising and artifact suppression has also been a successful endeavor. However, most of the published work have directly implemented the algorithms that have been originally proposed for natural images with little modification. It is likely that much better results could be achieved by designing dedicated algorithms for CT. In fact, CT images reconstructed from low-dose scans present unique challenges such as high noise levels with a nonuniform and unknown distribution and various types of artifacts. Although dealing with these images is challenging, the success of patch-based methods on natural images is a strong indication of their potential to tackle these challenges in CT.

In conclusion, this review shows that patch-based methods have a great potential in improving the current image reconstruction and processing algorithms in CT. With an ever increasing usage of CT in clinical applications, it is necessary to reduce the radiation dose used during imaging so that CT can be used to its full potential. Meanwhile, increased computational power makes it possible to use more sophisticated algorithms for image reconstruction and processing. Therefore, we believe that patch-based methods can play a key role in solving some of the major challenges facing CT.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This articles does not contain patient data.

References

1. Aharon M, Elad M (2008) Sparse and redundant modeling of image content using an image-signature-dictionary. SIAM J Imaging Sci 1(3):228–247
2. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322
3. AlAfeef A, Cockshott P, MacLaren I, McVitie S (2014) Compressed sensing electron tomography using adaptive dictionaries: a simulation study. J Phys: Conf Ser 522(1):012021
4. Bach F, Jenatton R, Mairal J, Obozinski G (2012) Optimization with sparsity-inducing penalties. Found Trends Mach Learn 4(1):1–106
5. Barnes C, Shechtman E, Goldman D, Finkelstein A (2010) The generalized patchmatch correspondence algorithm. In: Daniilidis K, Maragos P, Paragios N (eds) Computer vision ECCV 2010, vol 6313, lecture notes in computer ScienceSpringer, Berlin, pp 29–43
6. Borsdorf A, Köstler H, Rubinstein R, Bartuschat D, Strmer M (2009) A parallel K-SVD implementation for CT image denoising. Technical Report CS 10, University of Erlangen-Nurnberg
7. Bian Z, Ma J, Huang J, Zhang H, Niu S, Feng Q, Liang Z, Chen W (2013) Sr-nlm: a sinogram restoration induced non-local means

image filtering for low-dose computed tomography. Comput Med Imaging Graph 37(4):293–303
8. Buades A, Coll B, Morel J-M (2005) A review of image denoising algorithms, with a new one. Multiscale Model Simul 4(2):490–530
9. Buades A, Coll B, Morel J-M (2006) Image enhancement by non-local reverse heat equation. Preprint CMLA 22:2006
10. Chainais P (Sept 2012) Towards dictionary learning from images with non gaussian noise. In: 2012 IEEE international workshop on machine learning for signal processing (MLSP), pp 1–6
11. Chatterjee P, Milanfar P (2010) Is denoising dead? IEEE Trans Image Process 19(4):895–911
12. Chatterjee P, Milanfar P (2012) Patch-based near-optimal image denoising. IEEE Trans Image Process 21(4):1635–1649
13. Chen Y, Chen W, Yin X, Ye X, Bao X, Luo L, Feng Q, li Y, Yu X (2011) Improving low-dose abdominal CT images by weighted intensity averaging over large-scale neighborhoods. Eur J Radiol 80(2):e42–e49
14. Chen Y, Gao D, Nie C, Luo L, Chen W, Yin X, Lin Y (2009) Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior. Comput Med Imaging Graph 33(7):495–500
15. Chen Y, Shi L, Feng Q, Yang J, Shu H, Luo L, Coatrieux J-L, Chen W (2014) Artifact suppressed dictionary learning for low-dose CT image processing. IEEE Trans Med Imaging 33(12):2271–2292
16. Chen Y, Shi L, Yang J, Hu Y, Luo L, Yin X, Coatrieux J-L (2014) Radiation dose reduction with dictionary learning based processing for head CT. Australas Phys Eng Sci Med 37(3):483–493
17. Chen Y, Yang Z, Hu Y, Yang G, Zhu Y, Li Y, luo L, Chen W, Toumoulin C (2012) Thoracic low-dose CT image processing using an artifact suppressed large-scale nonlocal means. Phys Med Biol 57(9):2667
18. Chen Y, Yin X, Shi L, Shu H, Luo L, Coatrieux J-L, Toumoulin C (2013) Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing. Phys Med Biol 58(16):5803
19. Couzinie-Devy F, Mairal J, Bach F, Ponce J (2011) Dictionary learning for deblurring and digital zoom. Preprint. arXiv:1110.0957
20. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans Image Process 16(8):2080–2095
21. Darbon J, Cunha A, Chan T, Osher S, Jensen G (2008) Fast nonlocal filtering applied to electron cryomicroscopy. In: 5th IEEE international symposium on biomedical imaging: from Nano to Macro, 2008. ISBI 2008, pp 1331–1334, May
22. Deledalle C-A, Denis L, Tupin F (2012) How to compare noisy patches? patch similarity beyond gaussian noise. Int J Comput Vis 99(1):86–102
23. Deledalle C-A, Duval V, Salmon J (2012) Non-local methods with shape-adaptive patches (NLM-SAP). J Math Imaging Vis 43(2):103–120
24. Deledalle C-A, Salmon J, Dalalyan A (2011) Image denoising with patch based PCA: local versus global. BMVC 81:425–455
25. Deledalle C-A, Tupin F, Denis L (Sept 2010) Poisson NL means: unsupervised non local means for poisson noise. In: 17th IEEE international conference on image processing (ICIP), 2010, pp 801–804
26. Deledalle C-A, Tupin F, Denis L (Sept 2011) Patch similarity under non Gaussian noise. In: International conference on image processing. Brussels, pp 1845–1848
27. Dong W, Zhang D, Shi G, Wu X (2011) Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Trans Image Process 20(7):1838–1857
28. Dore V, Cheriet M (2009) Robust NL-means filter with optimal pixel-wise smoothing parameter for statistical image denoising. IEEE Trans Signal Process 57(5):1703–1716
29. Dupe F-X, Anthoine S (Sept 2013) A greedy approach to sparse poisson denoising. In: 2013 IEEE international workshop on machine learning for signal processing (MLSP), pp 1–6
30. Elad M (2010) Sparse and redundant representations: from theory to applications in signal and image processing. Springer, Berlin
31. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15(12):3736–3745
32. Engan K, Aase S, Hakon Husoy J (1999) Method of optimal directions for frame design. In 1999. Proceedings, 1999 IEEE international conference on acoustics, speech, and signal processing, vol 5, pp 2443–2446
33. Etter V, Jovanovic I, Vetterli M (2011) Use of learned dictionaries in tomographic reconstruction. In: SPIE optical engineering+ applications, International Society for Optics and Photonics, pp 81381C–81381C-11
34. Giryes R, Elad M (2013) Sparsity based poisson denoising with dictionary learning. Preprint. arXiv:1309.4306
35. Gregor K, Lecun Y (2010) Learning fast approximations of sparse coding. In: 2010 international conference on machine learning (ICML), Omnipress, pp 1–8
36. Hawe S, Seibert M, Kleinsteuber M (2013) Separable dictionary learning. In: 2013 IEEE conference on computer vision and pattern recognition (CVPR), IEEE, pp 438–445
37. Huang D-A, Wang Y-CF (2013) Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition. In: 2013 IEEE international conference on computer vision (ICCV), IEEE, pp 2496–2503
38. Huang J, Ma J, Liu N, Zhang H, Bian Z, Feng Y, Feng Q, Chen W (2011) Sparse angular CT reconstruction using non-local means based iterative-correction POCS. Comput Biol Med 41(4):195–205
39. Jenatton R, Mairal J, Bach FR, Obozinski GR (2010) Proximal methods for sparse hierarchical dictionary learning. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 487–494
40. Jia X, Tian Z, Lou Y, Sonke J-J, Jiang SB (2012) Four-dimensional cone beam CT reconstruction and enhancement using a temporal nonlocal means method. Med Phys 39(9):5592–5602
41. Kang L-W, Lin C-W, Fu Y-H (2012) Automatic single-image-based rain streaks removal via image decomposition. IEEE Trans Image Process 21(4):1742–1755
42. Karimi D, Ward RK (2016) A novel structured dictionary for fast processing of 3D medical images, with application to computed tomography restoration and denoising. In: SPIE medical imaging, International Society for Optics and Photonics, pp 97840N–97840N-7
43. Karimi D, Ward R (2016) Reducing streak artifacts in computed tomography via sparse representation in coupled dictionaries. Med Phys 43(3):1473–1486
44. Karimi D, Ward R, Ford N (Sept 2015) Angular upsampling of projection measurements in 3D computed tomography using a sparsity prior. In: 2015 IEEE international conference on image processing (ICIP), pp 3363–3367
45. Karimi D, Ward RK (2016) Sinogram denoising via simultaneous sparse representation in learned dictionaries. Phys Med Biol 61(9):3536–3553
46. Kavukcuoglu K, Ranzato M, Fergus R, Le-Cun Y (June 2009) Learning invariant features through topographic filter maps. In: IEEE conference on computer vision and pattern recognition, 2009. CVPR 2009, pp 1605–1612
47. Kazantsev D, Thompson WM, Van Eyndhoven G, Dobson K, Kaestner AP, Lionheart W, Withers PJ, Lee PD (2015) 4D-CT reconstruction with unified spatial-temporal patch-based regularization. Inverse Probl. Imaging 9:447–467
48. Kazantsev D, Van Eyndhoven G, Lionheart W, Withers P, Dobson K, McDonald S, Atwood R, Lee P (2015) Employing temporal

self-similarity across the entire time domain in computed tomography reconstruction. Philos Trans R Soc Lond A 373(2043):20140389
49. Kelm Z, Blezek D, Bartholmai B, Erickson B (June 2009) Optimizing non-local means for denoising low dose CT. In: IEEE international symposium on biomedical imaging: from Nano to Macro, 2009. ISBI '09, pp 662–665
50. Kindermann S, Osher S, Jones PW (2005) Deblurring and denoising of images by nonlocal functionals. Multiscale Model Simul 4(4):1091–1115
51. Kumar N, Zhang L, Nayar SK (Oct 2008) What is a good nearest neighbors algorithm for finding similar patches in images? In: European conference on computer vision (ECCV), pp 364–378
52. Lebrun M, Buades A, Morel JM (2013) A nonlocal Bayesian image denoising algorithm. SIAM J Imaging Sci 6(3):1665–1688
53. Lee H, Battle A, Raina R, Ng AY (2006) Efficient sparse coding algorithms. In: Advances in neural information processing systems, pp 801–808
54. Lewicki MS, Olshausen BA (1999) Probabilistic framework for the adaptation and comparison of image codes. JOSA A 16(7):1587–1601
55. Li S, Cao Q, Chen Y, Hu Y, Luo L, Toumoulin C (2014) Dictionary learning based sinogram inpainting for CT sparse reconstruction. Optik Int J Light Electron Optics 125(12):2862–2867
56. Li S, Fang L, Yin H (2012) An efficient dictionary learning algorithm and its application to 3-D medical image denoising. IEEE Trans Biomed Eng 59(2):417–427
57. Li Z, Yu L, Trzasko JD, Lake DS, Blezek DJ, Fletcher JG, McCollough CH, Manduca A (2014) Adaptive nonlocal means filtering based on local noise level for CT denoising. Med Phys 41(1):011908
58. Liu B, Yu H, Verbridge SS, Sun L, Wang G (2014) Dictionary-learning-based reconstruction method for electron tomography. Scanning 36(4):377–383
59. Lou Y, Zhang X, Osher S, Bertozzi A (2010) Image recovery via nonlocal operators. J Sci Comput 42(2):185–197
60. Lu Y, Zhao J, Wang G (2012) Few-view image reconstruction with dual dictionaries. Phys Med Biol 57(1):173
61. Ma J, Huang J, Feng Q, Zhang H, Lu H, Liang Z, Chen W (2011) Low-dose computed tomography image restoration using previous normal-dose scan. Med Phys 38(10):5713–5731
62. Ma J, Zhang H, Gao Y, Huang J, Liang Z, Feng Q, Chen W (2012) Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior. Phys Med Biol 57(22):7519
63. Maggioni M, Katkovnik V, Egiazarian K, Foi A (2013) A nonlocal transform-domain filter for volumetric data denoising and reconstruction. IEEE Trans Image Process 22(1):1057–7149
64. Mairal J, Bach F, Ponce J (2012) Task-driven dictionary learning. IEEE Trans Pattern Anal Mach Intell (PAMI) 34(4):791–804
65. Mairal J, Bach F, Ponce J (2014) Sparse modeling for image and vision processing. Found Trends Comput Graph Vis 8(2–3):85–283
66. Mairal J, Bach F, Ponce J, Sapiro G (2010) Online learning for matrix factorization and sparse coding. J Mach Learn Res (JMLR) 11:19–60
67. Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A (2009) Non-local sparse models for image restoration. In: International conference on computer vision (ICCV). I
68. Mairal J, Sapiro G, Elad M (2008) Learning multiscale sparse representations for image and video restoration. Multiscale Model Simul 7(1):214–241
69. Peyr G, Bougleux S, Cohen L (2008) Non-local regularization of inverse problems. In: Forsyth D, Torr P, Zisserman A (eds) Computer vision ECCV 2008, vol 5304, Lecture notes in computer ScienceSpringer, Berlin, pp 57–68
70. Pfister L, Bresler Y (2014) Model-based iterative tomographic reconstruction with adaptive sparsifying transforms . In: IS&T/SPIE electronic imaging, International Society for Optics and Photonics, pp 90200H–90200H
71. Protter M, Elad M, Takeda H, Milanfar P (2009) Generalizing the nonlocal-means to super-resolution reconstruction. IEEE Trans Image Process 18(1):36–51
72. Ramirez I, Sprechmann P, Sapiro G (June 2010) Classification and clustering via dictionary learning with structured incoherence and shared features. In: 2010 IEEE conference on computer vision and pattern recognition (CVPR), pp 3501–3508
73. Rubinstein R, Zibulevsky M, Elad M (2010) Double sparsity: learning sparse dictionaries for sparse signal approximation. IEEE Trans Signal Process 58(3):1553–1564
74. Sakhaee E, Entezari A (2014) Learning splines for sparse tomographic reconstruction. In: Bebis G, Boyle R, Parvin B, Koracin D, McMahan R, Jerald J, Zhang H, Drucker S, Kambhamettu C, El Choubassi M, Deng Z, Carlson M (eds) Advances in visual computing, vol 8887. Lecture notes in computer science. Springer International Publishing, pp 1–10
75. Salmon J, Harmany Z, Deledalle C-A, Willett R (2014) Poisson noise reduction with non-local PCA. J Math Imaging Vis 48(2):279–294
76. Shtok J, Elad M, Zibulevsky M (2013) Learned shrinkage approach for low-dose reconstruction in computed tomography. J Biomed Imaging 2013:7
77. Shtok J, Elad P, Zibulevsky M (May 2011) Sparsity-based sinogram denoising for low-dose computed tomography. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 569–572
78. Soltani S (2015) Studies of sensitivity in the dictionary learning approach to computed tomography: simplifying the reconstruction problem, rotation, and scale. DTU Compute-Technical Report-2015. Technical University of Denmark
79. Soltani S, Andersen MS, Hansen PC (2015) Tomographic image reconstruction using dictionary priors. CoRR, abs/1503.01993
80. Stojanovic I, Pien H, Do S, Karl W (May 2012) Low-dose X-ray CT reconstruction based on joint sinogram smoothing and learned dictionary-based representation. In: 2012 9th IEEE international symposium on biomedical imaging (ISBI), pp 1012–1015
81. Tan S, Zhang Y, Wang G, Mou X, Cao G, Wu Z, Yu H (2015) Tensor-based dictionary learning for dynamic tomographic reconstruction. Phys Med Biol 60(7):2803
82. Tian Z, Jia X, Dong B, Lou Y, Jiang SB (2011) Low-dose 4DCT reconstruction via temporal nonlocal means. Med Phys 38(3):1359–1365
83. Wang S, Zhang L, Liang Y, Pan Q (2012) Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In 2012 IEEE conference on computer vision and pattern recognition (CVPR), IEEE, pp 2216–2223
84. Xu Q, Yu H, Mou X, Zhang L, Hsieh J, Wang G (2012) Low-dose X-ray CT reconstruction via dictionary learning. IEEE Trans Med Imaging 31(9):1682–1697
85. Xu W, Ha S, Mueller K (2013) Database-assisted low-dose CT image restoration. Med Phys 40(3):031109
86. Xu W, Mueller K (2012) Efficient low-dose CT artifact mitigation using an artifact-matched prior scan. Med Phys 39(8):4748–4760
87. Yan H, Zhen X, Cervio L, Jiang SB, Jia X (2013) Progressive cone beam CT dose control in image-guided radiation therapy. Med Phys 40(6):060701
88. Yang J, Wright J, Huang T, Ma Y (2010) Image super-resolution via sparse representation. IEEE Trans Image Process 19(11):2861–2873
89. Zhang H, Huang J, Ma J, Bian Z, Feng Q, Lu H, Liang Z, Chen W (2014) Iterative reconstruction for X-ray computed tomogra-

phy using prior-image induced nonlocal regularization. IEEE Trans Biomed Eng 61(9):2367–2378
90. Zhang H, Ma J, Wang J, Liu Y, Lu H, Liang Z (2014) Statistical image reconstruction for low-dose CT using nonlocal means-based regularization. Comput Med Imaging Graph 38(6):423–435
91. Zhang X, Burger M, Bresson X, Osher S (2010) Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM J Imaging Sci 3(3):253–276
92. Zhao B, Ding H, Lu Y, Wang G, Zhao J, Molloi S (2012) Dual-dictionary learning-based iterative image reconstruction for spectral computed tomography application. Phys Med Biol 57(24):8217
93. Zhou M, Chen H, Paisley J, Ren L, Sapiro G, Carin L (2009) Non-parametric Bayesian dictionary learning for sparse image representations 1. In: NIPS 2009