Iterative CT reconstruction using shearlet-based regularization

Bert Vandeghinste, Bart Goossens, Roel Van Holen, *Member, IEEE*, Christian Vanhove, Aleksandra Pižurica, Stefaan Vandenberghe, and Steven Staelens

Abstract-Total variation (TV) methods have been proposed to improve the image quality in count-reduced images, by reducing the variation between neighboring pixels. Although very easy to implement and fast to compute, TV-based methods may lead to a loss of texture information when applied to images with complex textures, such as high-resolution abdominal CT images. Here, we investigate the use of another regularization approach in the context of medical images based on multiresolution transformations. One such transformation is the shearlet transform, which is optimally sparse for images that are C^2 except for discontinuities along C^2 curves, and has better directional sensitivity than most other, related, wavelet transform approaches. We propose to solve the convex problem using the split-Bregman (augmented Lagrangian) approach. One of the primary advantages of the split-Bregman approach, is that the shearlet transform can easily be incorporated into the sparse-view CT reconstruction.

The required sparsity prior is the ℓ_1 norm of the shearlet coefficients. Results are shown for this method in comparison to the same framework with TV as the regularization term on simulated data. The noise-resolution performance is investigated at different contrast levels. At equal image noise, TV-based regularization outperforms shearlet-based regularization. However, when image texture is analyzed on *measured* mouse data, shearlets outperform TV, which suffers from staircasing effects.

Our results show that there are benefits in using shearlets in CT imaging: texture is reconstructed more accurately compared to when TV is used, without biasing the image towards a piecewise constant image model. However, due to the larger support of the basis functions, our results suggest that uncareful usage of shearlets may lead to wavy artifacts, which can be equally unwanted as staircasing effects.

Index Terms—Computed Tomography (CT) reconstruction, Image reconstruction

I. INTRODUCTION

TOTAL variation (TV) minimization or regularization is one of the techniques that have been extensively investigated in the context of image denoising by different image processing groups [1]. In the context of compressed sensing, TV minimization has already been used for few-view, limited-angle CT reconstruction [2]–[6], interior tomography [7], [8] and CT image denoising or restoration [9], [10].

B. Vandeghinste, R. Van Holen, C. Vanhove and S. Vandenberghe are with MEDISIP, Department of Electronics and Information Systems, Ghent University–iMinds–IBiTech, De Pintelaan 185 Block B, B-9000 Ghent, Belgium e-mail: (Bert.Vandeghinste@UGent.be)

B. Goossens and A. Pižurica are with IPI, Department of Telecommunications and Information Processing, Ghent University–iMinds, Ghent, Belgium. S. Staelens is with the Molecular Imaging Center Antwerp, University of

Antwerp, 2650 Edegem, Belgium.

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. These regularizing methods in general take advantage of the natural redundancy present in images. CT images can be conceptualized as being the sum of edges, uniform intensities, a repeated structure of fine-scale patterns (texture) and a noise component [11], [12]. In image denoising applications it would then be beneficial if the representation of that image separates the structural components from the noise. This is only possible if the noise-free image can be approximated with a small number of significant coefficients in that representation. In that case, the representation is called a sparse representation. With TV regularization, the gradient operator is used as the sparsifying transformation, transforming the original image into an edge map.

TV-based methods are best suited for piece-wise constant images, such as images of simple geometric shapes with flat intensity. Some authors have reconstructed such a simple phantom (e.g. Shepp-Logan phantom [13], left on Fig. 1) exactly from only 22 projections [2], [3], [14], which is impressive and promising for dose-reduction applications in CT. However, TV minimization can not reconstruct an image from 82 realistic projections anymore [5], [15] when the simple phantom is replaced by a phantom specifically designed for this purpose (middle of Fig. 1). Some medically relevant information, such as a tumor in the brain, is inevitably lost, because the tumor was designed to be invisible for a certain number of projection views (ghosting [5], [15]). When the phantoms or images become even more complex, i.e. images that contain complex textures and gradual intensity transitions like high-resolution lung CT images (e.g. right image on Fig. 1, from [16]), TV based methods often produce cartoon-like approximations. At least 100 views are necessary when realistic human anatomy is present [17]. The same holds for μ CT [18] as well as for MRI [19], [20]. This discrepancy between results obtained from simple phantoms and from realistic data reflects the fact that most natural images are not of bounded variation [21], [22], the search space in which TV minimization operates. Only the edge component can generally be expected to be of bounded variation [21]. While total variation minimization keeps the noise component small, the texture component is often eliminated due to the staircasing effect [23]-[26]. Unfortunately, high-resolution medical images are not of bounded variation [5], [21]. This makes the TV regularizer unsuited for medical images from a mathematical point of view. Consequently, it is beneficial to investigate regularizers that offer a better sparse representation of medical images than TV.

Several alternatives have been suggested in the literature.



Fig. 1. Left: Shepp-Logan phantom [13], which only requires 22 projection views to reconstruct completely. Middle: More complex phantom, as used in [5], [15]. Right: Addition of realistic lung texture to the XCAT phantom [16], [27] (Window [-900 0] HU).

One possibility is the use of the wavelet transformation. This transformation provides a decomposition of a signal over dilated and translated versions of a fixed waveform, called the mother wavelet. It allows the regularization to adapt to the image content at different resolution scales. While TV minimization might destroy soft edges, these would usually be reconstructed better using wavelets. The wavelet coefficients evolve across the scales at a rate that depends on the local regularity of the signal (e.g. depending on edge smoothness). Hence, even very "weak" edges produce a significant response at a certain scale. In this way, the wavelet coefficients give valuable information about the edge which the discrete gradient operator lacks.

The Haar wavelet, the simplest possible wavelet, has already been investigated for regularized CT reconstruction by Garduño et al. [15]. They found that reconstructions with a sparse Haar transform are not more effective from the medical diagnostic point of view than reconstructions that have a small TV value. The authors did not explain this result. Thus, the search for an objective function that provides diagnostically efficacious reconstruction from a limited number of CT projections remained open [15].

However, Haar wavelets are optimal for piecewise constant images and are very similar to TV in this regard. Steidl et al. [28] have proven that for 1-D signals, Haar wavelet shrinkage is equivalent to a single step of space-discrete TV diffusion or regularization of two pixel-pairs, when applied to one single scale only. To our knowledge the higher-dimensional case has not been proven yet, since it cannot be treated as a straightforward generalization of 1-D ideas [28]. The near equivalence of the Haar-based regularization to the TV regularization in 1-D might explain why Haar-based regularization did not outperform TV regularization.

Furthermore, it is known that the wavelet transform in general has poor directional selectivity. The discrete wavelet transform is often computed by using basis functions that are the tensor product of one-dimensional wavelets and onedimensional scaling functions. This construction of wavelets will produce a checkerboard pattern simultaneously oriented along several directions [29], also known as the checkerboard problem. This approach can deal with point-wise singularities (such as point sources), but does not allow to make a distinction between features at $+45^{\circ}$ and -45° . As a result, many nonzero wavelet coefficients may be needed to represent a line singularity at an arbitrary orientation.

The poor orientation selectivity of the wavelet transformation led to the development of a number of multiresolution geometrical transformations with typically a very high number of analysis orientations, such as ridgelets [30], [31] and curvelets [32], [33]. The ridgelet transform is well suited for representing discontinuities along straight lines, in contrast to the curvelet transform, which can even represent discontinuities along curves with bounded curvatures. A recent development is the shearlet transform [34]. This non-isotropic version of the wavelet transform is comparable to curvelets, as it also performs multiscale and multidirectional analysis, and both transformations can represent curve-like singularities in images. Guo et al. [35] have shown that the asymptotic decay rate of the shearlet transform, for fine scales, can be used to signal both the location and the orientation of the edges of an image and that the coefficients of large magnitude will correspond to edges. Furthermore, the decay rate across scales can be used to distinguish between noise spikes and edges [36], a property also holding for wavelet coefficients.

There are a number of advantages in using shearlets in imaging [37], [38]. The primary advantage for our use case is that shearlets allow for a lower redundant sparse tight frame representation than other related multiresolution representations (such as e.g., ridgelets, curvelets, ...), while still offering shift invariance and a directional analysis. Basically, the number of transform coefficients is not much larger than the number of pixels in the original image, and no artifacts are introduced when shearlet coefficients are adjusted. The directional analysis allows the shearlet basis functions to align with x-ray noise streaks in the reconstructed images, which allows these streaks to be approximated with less significant coefficients.

Secondly, the shearlet representation can be used to decompose the space $L^2(\mathbb{R}^2)$ of images into a sequence of spaces, associated to a hierarchy of scales [39], [40]. While edges are quantified in the space of bounded variation $BV(\mathbb{R}^2)$, other relevant features such as homogenous regions, texture, and other oscillatory patterns will belong to spaces between the smaller $BV(\mathbb{R}^2)$ and the larger $L^2(\mathbb{R}^2)$ [40]. By setting shearlet coefficients below a certain threshold to zero, image features present above a certain scale can be extracted [39].

Both properties are useful when we apply soft thresholding (soft shrinkage) later on in Section II-2. A third property is that shearlets take edges into account in a multitude of directions, decomposed into a hierarchy of scales. This allows the detection of soft edges. Lastly, in compressed sensing the goal is to use a transformation with the best approximation (compaction) properties. In higher dimensions, shearlets are a better candidate for this than wavelets or TV, as shearlets have an essentially optimal approximation error for images that contain edges (images that are C^2 (i.e. twice continuously differentiable) apart from discontinuities along C^2 curves) [34].

Shearlets are very similar to curvelets, as both perform a multiscale and multidirectional analysis. In fact, both trans-

forms have optimal sparsity for images which have discontinuities (edges) along a C^2 curve [34], [41]. However, some important differences remain [38]. Shearlets are generated by applying a family of operators to a single function, contrary to curvelet basis functions. This makes the discrete implementation of curvelets very challenging. Two methods were suggested by Candès et al. [32] to make the implementation easier. In a first method, aliasing is deliberately introduced through wrapping. In a second method, the nonequispaced fast Fourier transform (NFFT) is used. The inverse NFFT should then be computed with a conjugate gradient type algorithm, utilizing one NFFT and one adjoint NFFT per iteration [42]. This makes the calculations less efficient. A second difference with curvelets is that shearlets are associated to a multiresolution analysis, while curvelets are not. Thirdly, in the construction of the shearlet tight frame, the number of orientations doubles at every scale, while for curvelets, this number doubles every other scale. A final difference, and very important one, is that shearlets allow for a much less redundant sparse tight frame representation than curvelets, while still offering shift invariance [37]. These properties make the shearlet transform an attractive candidate for image representation [37].

In this paper, we investigate if shearlets can be used in regularized CT reconstruction, and if they show any of the previously mentioned benefits compared to TV regularization. This is based on the results of shearlet-based denoising in image processing [38], [39], in magnetic resonance imaging (MRI) reconstruction [43] and preliminary numeric results we published previously [44]. Previously, Colonna et al. [45] have used the shearlet representation to invert the Radon transform directly. Their approach is not directly applicable to ill-posed systems, i.e. sparse viewing. We propose to use a recently proposed efficient solver, based on an augmented Lagrangian (or split-Bregman) approach [46], [47], as an alternative method for sparse-view CT reconstruction. Split-Bregman methods have been found to be successful in other applications, such as MRI image reconstruction [43]. This very general method allows the incorporation of extra regularization terms quite easily.

II. THE SPLIT-BREGMAN BASED CT RECONSTRUCTION METHOD

Imaging model

In all the following, we will denote an ℓ_1 -norm by $|.|_1$, an ℓ_2 -norm by $||.||_2$. Matrices are in bold uppercase font, vectors are typeset in a bold lowercase font. In a noise free case, the projection data y can be modeled by the discrete approximation of the imaging process $\mathbf{y} = \mathbf{W}\mathbf{x}$, with W the system matrix modeling the X-ray transformation and x the reconstructed image.

With no noise present and an underdetermined system (less projection samples than pixels in the image), solving for x is defined as minimizing the least-squares cost function:

$$g(\mathbf{x}) = \|\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2.$$
(1)

The optimal solution $\hat{\mathbf{x}}$ that minimizes (1) is then given by the pseudoinverse of \mathbf{W} . Unfortunately, this pseudoinverse is

too complicated to compute directly in practical CT imaging. An alternative, iterative, method to find \mathbf{x} is using gradient descent steps $\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} = -2\mathbf{W}^T (\mathbf{y} - \mathbf{W}\mathbf{x})$, such that

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + 2\lambda \mathbf{W}^T \left(\mathbf{y} - \mathbf{W} \mathbf{x}^{(i)} \right), \qquad (2)$$

with *i* the iteration number and λ a relaxation parameter. Assuming that the row and column sums of **W** are equal to 1, λ equals 1/2 and correct scaling is applied, this gives the classical Simultaneous Iterative Reconstruction Technique (SIRT) algorithm [48], [49]. This gradient descent algorithm will converge to the same solution $\hat{\mathbf{x}}$ as would be obtained if the pseudoinverse would be calculated directly.

In the presence of noise in projection space, the system can be modeled by an additive approximation $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$, with **n** given by a Gaussian Random Field [37], [50]. Because the noise is zero-mean by approximation, the data fitting function (1) is still applicable.

Minimizing (1) for x is difficult with the addition of noise, and even more difficult in the case of sparse-view systems. This makes the problem very ill-posed. Therefore, we define a new cost criterium with an added regularization term:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{arg\,min}} \ |\Phi(\mathbf{x})|_1 \quad \text{s.t.} \quad \left\| \mathbf{C}^{-1} \left(\mathbf{y} - \mathbf{W} \mathbf{x} \right) \right\|_2^2 \le \epsilon.$$
 (3)

Here, ϵ acts as an upper bound on the uncertainty about the measurements y, and Φ is a linear sparsifying transformation.

Matrix C is in first instance applied to weight the errors corresponding to projections with high attenuation (high values of y) [51]. These detected values will contain more noise than values projected through less dense material, and will thus have lower signal-to-noise ratio (SNR). Therefore, we need to weight the projected values according to their variance, which is equal to the mean projection value under the assumption of poisson noise:

$$\mathbf{C} = \text{diag}(c_0, c_1, ..., c_i) \quad \text{with} \quad c_i = e^{-\mathbf{E}[y_i]}.$$
 (4)

We can use the projection value itself as an estimate for the expected value $E[y_i]$.

Next to simple exponential weighing, C can also be used to model the detector acquisition system. Different effects can be taken into account in a diagonal matrix, such as gain per detector pixel, thermal stability, dark current drifts and linearity [52]. These effects can be modeled by multiplying c_i with appropriately measured air scans. Signal crosstalk can be modeled with off-diagonal covariance entries [53], as well as afterglow and primary speed. In that case, C serves as a prewhitener for this system, decorrelating the noise.

Because Eq. (3) is non-differentiable (the ℓ_1 -norm has a discontinuity at the origin), it should be converted to an unconstrained problem [47]:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{arg\,min}} \left\| \Phi(\mathbf{x}) \right\|_{1} + \lambda \left\| \mathbf{C}^{-1} \left(\mathbf{y} - \mathbf{W} \mathbf{x} \right) \right\|_{2}^{2}, \qquad (5)$$

with λ a constant. In words, among all solutions which are closest (in the weighted least-squares sense) to the acquired data, equation (5) finds a solution which is sparse in the ℓ_1 -sense in the domain of the transform Φ . The use of the ℓ_1 -norm leads to suppression of many small coefficients in favor

of a few large coefficients, and is thus crucial to the whole approach of regularization with sparsifying transforms [19].

Overview of the split-Bregman framework

There is a vast amount of literature available on how to solve equations such as (5) in general, and in particular adapted for CT reconstruction. In the majority of cases it is solved through a heuristical method, alternating the optimization of the data fitting constraint with the sparsity constraint [2], [4], [6], [7], [54]. We will use the split-Bregman framework [47] (also known as the split-augmented Lagrangian). This is very similar to work done in iterative thresholding [55], [56], with the benefits of having a relatively low memory footprint [47], simple and fast iteration steps, and that the technique is generally easy to implement, even for complex problems.

Goldstein and Osher [47] have shown that the generalized constrained optimization problem

$$\operatorname*{arg\,min}_{\mathbf{x}} J(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{y} = \mathbf{W} \mathbf{x} \tag{6}$$

with J a convex energy functional, can be solved by iterating over

$$\mathbf{x}^{(i+1)} = \operatorname*{arg\,min}_{\mathbf{x}} J(\mathbf{x}) + \frac{\lambda}{2} \left\| \mathbf{W} \mathbf{x} - \mathbf{b}^{(i)} \right\|_{2}^{2} \qquad (7a)$$

$$\mathbf{b}^{(i+1)} = \mathbf{b}^{(i)} + \mathbf{y} - \mathbf{W}\mathbf{x}^{(i)}.$$
 (7b)

The iterates $\mathbf{x}^{(i)}$ will get arbitrarily close to a solution of the original constrained problem (6).

Applying (7a) and (7b) to cost function (5) together with variable splitting ultimately leads to the following three update equations [47], [57]:

$$\mathbf{x}^{(i+1)} = \underset{\mathbf{x}}{\arg\min} \frac{\lambda}{2} \left\| \mathbf{C}^{-1} \left(\mathbf{y} - \mathbf{W} \mathbf{x} \right) \right\|_{2}^{2}$$

$$+ \frac{\mu}{2} \left\| \mathbf{d}^{(i)} - \Phi(\mathbf{x}) - \mathbf{b}^{(i)} \right\|_{2}^{2}$$
(8a)

$$\mathbf{d}^{(i+1)} = \underset{\mathbf{d}}{\operatorname{arg\,min}} \left\| \mathbf{d} \right\|_{1}$$

$$+ \frac{\mu}{2} \left\| \mathbf{d} - \Phi \left(\mathbf{x}^{(i+1)} \right) - \mathbf{b}^{(i)} \right\|_{2}^{2}$$
(8b)

$$\mathbf{b}^{(i+1)} = \mathbf{b}^{(i)} + \left(\Phi\left(\mathbf{x}^{(i+1)}\right) - \mathbf{d}^{(i+1)}\right).$$
(8c)

The ℓ_1 - and ℓ_2 -norm from the regularized quadratic problem (5) are now split into different minimization problems: a sequence of unconstrained optimization problems and one Bregman update step. Eq. (8b) and (8c) are easy to solve, and Eq. (8a) only needs to be solved for $\mathbf{x}^{(i+1)}$ approximately. Even then, the algorithm still converges [47].

We have introduced two new variables which elicit further explanation: λ and μ .

Parameter λ was added by the conversion of the constrained optimization problem to an unconstrained one. In this context, it can be interpreted as the contribution of the regularization to the total cost. The lower its value, the lower the importance of the data fitting term, which amounts to more regularization (denoising).

The ratio μ/λ determines the convergence speed. If μ is set to 0 and λ to 1, no regularization is performed and the classical SIRT algorithm is obtained, minimizing cost function

(1). A ratio of 2.0 was empirically determined to deliver a good convergence speed [47]. We will experimentally determine good parameters for our study further on in Section V-B.

Regularization terms

In the previous sections Φ was intentionally left unspecified. Indeed, no assumptions have been made apart from convexity. The purpose of this study is to investigate if using the discrete shearlet transformation as a regularization term is superior to isotropic TV minimization. Isotropic TV was chosen as this is what is used most widely in medical imaging.

1) Isotropic Total Variation: For TV minimization the discrete gradient operator ∇ can be used as Φ . For 2D isotropic TV, $J(\mathbf{x})$ then becomes:

$$J(\mathbf{x}) = \sum \sqrt{\left(\nabla_x \left(\mathbf{x}\right)\right)^2 + \left(\nabla_y \left(\mathbf{x}\right)\right)^2}.$$
 (9)

The minimization problem can then be solved by setting $\mathbf{d}_x = \nabla_x (\mathbf{x})$ and $\mathbf{d}_y = \nabla_y (\mathbf{x})$ [47]. Despite the fact that \mathbf{d}_x and \mathbf{d}_y can not be decoupled, the problem can still be solved using a generalized shrinkage formula [47], [58].

Equations (8a-8c) become:

$$\mathbf{x}^{(i+1)} = \arg\min_{\mathbf{x}} \frac{\lambda}{2} \| \mathbf{C}^{-1} (\mathbf{y} - \mathbf{W}\mathbf{x}) \|_{2}^{2}$$
(10a)
+ $\frac{\mu}{2} \| \mathbf{d}_{x}^{(i)} - \nabla_{x}(\mathbf{x}) - \mathbf{b}_{x}^{(i)} \|_{2}^{2}$
+ $\frac{\mu}{2} \| \mathbf{d}_{y}^{(i)} - \nabla_{y}(\mathbf{x}) - \mathbf{b}_{y}^{(i)} \|_{2}^{2}$
$$\mathbf{s}^{(i+1)} = \sqrt{\left| \nabla_{x} (\mathbf{x}^{(i+1)}) + \mathbf{b}_{x}^{(i)} \right|^{2} + \left| \nabla_{y} (\mathbf{x}^{(i+1)}) + \mathbf{b}_{y}^{(i)} \right|^{2}}$$
(10b)

$$\mathbf{d}_{\{x,y\}}^{(i+1)} = \max\left(\mathbf{s}^{(i+1)} - 1/\mu, 0\right) \frac{\nabla_{\{x,y\}}\left(\mathbf{x}^{(i+1)}\right) + \mathbf{b}_{\{x,y\}}^{(i+1)}}{\mathbf{s}^{(i+1)}}$$
(10c)

$$\mathbf{b}_{\{x,y\}}^{(i+1)} = \mathbf{b}_{\{x,y\}}^{(i)} + \left(\nabla_{\{x,y\}}\left(\mathbf{x}^{(i+1)}\right) - \mathbf{d}_{\{x,y\}}^{(i+1)}\right).$$
(10d)

The discrete gradient operator $\nabla_{\{x,y\}}(\mathbf{x})$ is defined as:

$$\nabla_x(\mathbf{x})(i,j) = \mathbf{x}(i,j) - \mathbf{x}(i-1,j)$$
(11a)

$$\nabla_y(\mathbf{x})(i,j) = \mathbf{x}(i,j) - \mathbf{x}(i,j-1), \quad (11b)$$

with wrapping of the values at the edges of the image volume.

2) Discrete shearlet transformation: The main mathematical ideas of the discrete shearlet transformation (DST) are reiterated here for completeness. For more detail, the reader is referred to the literature [34], [38], [59], [60].

Let $\psi_{j,k,1}(\mathbf{x})$ denote the shearlet basis functions (or in the remainder simply called shearlets), then the continuous shearlet transformation (CST) of an image $f(\mathbf{x}) \in L_2(\mathbb{R}^2)$ is defined by [35], [61]:

$$[\mathcal{SH}_{\psi}f](j,k,\mathbf{l}) = \int_{\mathbb{R}^2} f(\mathbf{x})\psi_{j,k,\mathbf{l}}(\mathbf{l}-\mathbf{x})dx \qquad (12)$$

where $j \in \mathbb{R}$, $k \in \mathbb{R}$ and $\mathbf{l} \in \mathbb{R}^2$ denote the scale, orientation and the spatial location, respectively. Shearlets are formed by dilating, shearing and translating a mother shearlet function $\psi \in L_2(\mathbb{R}^2)$, as follows:

$$\psi_{j,k,\mathbf{l}}(\mathbf{x}) = |\det \mathbf{A}|^{j/2} \psi \left(\mathbf{B}^k \mathbf{A}^j \mathbf{x} - \mathbf{l} \right)$$
(13)

where A and B are invertible 2×2 matrices, with det B = 1.

For shearlet analysis, the transformation matrices $\mathbf{A} = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ are used. A is an anisotropic scaling matrix with scaling factor a > 0 and **B** is a geometric shear matrix with parameter $s \in \mathbb{R}$.

The shearlet mother function ψ is a composite wavelet that satisfies appropriate admissibility conditions [35]. It is defined in the Fourier transform domain by $\Psi(\omega) =$ $\Psi_1(\omega_x) \Psi_2\left(\frac{\omega_y}{\omega_x}\right)$, with $\omega = [\omega_x \omega_y]$, $\Psi_1(\omega_x)$ the Fourier transform of a wavelet function and $\Psi_2(\omega_y)$ a compactly supported bump function $\Psi_2(\omega_y) = 0 \Leftrightarrow \omega_y \notin [-1, 1]$.

Any $f \in L_2(\mathbb{R}^2)$ can be recovered via the reproducing formula:

$$f = \sum_{j,k,\mathbf{l},d} \langle f, \psi_{j,k,\mathbf{l}} \rangle \psi_{j,k,\mathbf{l}}.$$
 (14)

The DST was implemented as proposed by Goossens et al. [62]–[64], which leads to a very low redundancy factor and a relatively short computation time compared to other implementations. We implement Φ as the forward shearlet transformation (Eq. 12) and its adjoint Φ^{\dagger} as the inverse shearlet transformation (Eq. 14).

The Meyer wavelet is used as mother wavelet $\Psi_1(\omega_x)$ for the shearlet transformation. This mother wavelet is an appealing choice due to its excellent localization properties in both time and frequency and also because the filters are defined directly in the frequency domain [62], [65]. The angular filter used is given by:

$$\Psi_{2}(\omega) = \begin{cases} 0 & \omega < -\frac{1+\alpha}{2} \\ \sin\left(\frac{\pi}{2}v\left(\frac{\alpha+2\omega+1}{2\alpha}\right)\right) & \left|\omega+\frac{1}{2}\right| \le \frac{\alpha}{2} \\ 1 & \left|\omega\right| < \frac{1-\alpha}{2} \\ \cos\left(\frac{\pi}{2}v\left(\frac{\alpha+2\omega-1}{2\alpha}\right)\right) & \left|\omega-\frac{1}{2}\right| \le \frac{\alpha}{2} \\ 0 & \text{else} \end{cases}$$
(15)



Fig. 2. Plot showing shearlets for different values of the parameter α . The shearlets become more elongated with decreasing α values.

The choice of $\alpha \in [0, \frac{1}{2}]$ here is important, it is a constant parameter that determines the bandwidth of the angular filters (see Fig. 2). The choice of α also has an influence on the redundancy factor of the DST.

Figure 3 shows all subbands of the shearlet transformation using 3 scales with 4 directions per scale, applied to measured



Fig. 3. Example of shearlet transformation coefficients, using 3 subbands with 4 directions per subband and $\alpha = 1/2$, applied to measured plastimouse data. All coefficient images show the magnitude of the coefficients (white equals a high magnitude, black equals zero. Magnitude normalized per subband). Images are for illustrative purposes and not to scale.

plastimouse data (see Section III-B). For an input image of 512×512 voxels, this leads to subbands with respective sizes 16×16 , 64×64 , 256×256 and 512×512 coefficients.

CT reconstruction algorithm

The performance of the split-Bregman method is largely dependent on the subproblem solvers for Eq. (8a) and Eq. (8b). To find a solution to Eq. (8a), we search for the roots of its derivative:

$$\mathbf{W}^{\dagger}\mathbf{C}^{-1}(\mathbf{y} - \mathbf{W}\mathbf{x}) + \frac{\mu}{\lambda}\Phi^{\dagger}\left(\mathbf{d}^{(i)} - \Phi(\mathbf{x}) - \mathbf{b}^{(i)}\right) = 0.$$
(16)

This can be rewritten as

$$\left(\mathbf{W}^{\dagger}\mathbf{C}^{-1}\mathbf{W} + \frac{\mu}{\lambda}\Phi^{\dagger}\Phi\right)\mathbf{x} = \mathbf{W}^{\dagger}\mathbf{C}^{-1}\mathbf{y} + \frac{\mu}{\lambda}\Phi^{\dagger}\left(\mathbf{d}^{(i)} - \mathbf{b}^{(i)}\right)$$
(17)

Remember that \mathbf{W} is the forward projector and Φ is the forward sparsifying operator. We now also need \mathbf{W}^{\dagger} , the back projector, and Φ^{\dagger} , the backward sparsifying operator.

When shearlets are used, $\Phi^{\dagger}\Phi = \mathbf{I}$, because of the property that shearlets form a tight frame [38], [63]. However, in CT reconstructions $\mathbf{W}^{\dagger}\mathbf{C}^{-1}\mathbf{W}$ will usually not be equal to \mathbf{I} . \mathbf{C} may have off-diagonal elements due to modeling of e.g. detector element crosstalk or afterglow, and \mathbf{W} is not always square. Furthermore, applying \mathbf{W}^{\dagger} (direct backprojection) after \mathbf{W} will result in a blur proportional to 1/r with r the distance from the source. Thus for shearlets, Eq. (17) can not be easily solved through inversion. The same property holds when TV is to be used. For isotropic TV, Eq. (17) becomes:

$$\begin{pmatrix} \mathbf{W}^{\dagger} \mathbf{C}^{-1} \mathbf{W} + \frac{\mu}{\lambda} \left(\nabla_{x}^{\dagger} \nabla_{x} + \nabla_{y}^{\dagger} \nabla_{y} \right) \mathbf{x} = \mathbf{W}^{\dagger} \mathbf{C}^{-1} \mathbf{y} \\ + \frac{\mu}{\lambda} \left(\nabla_{x}^{\dagger} \left(\mathbf{d}_{x}^{(i)} - \mathbf{b}_{x}^{(i)} \right) + \nabla_{y}^{\dagger} \left(\mathbf{d}_{y}^{(i)} - \mathbf{b}_{y}^{(i)} \right) \right)$$
(18)

We remark that $\nabla_x^{\dagger} \nabla_x + \nabla_y^{\dagger} \nabla_y \neq \mathbf{I}$. Because of these issues for both shearlets and TV, we selected the conjugate gradient (CG) method [66] to find a solution for Eq. (17). In Eq. (8b), there is no coupling between elements of d. The optimal value of d can then be computed using the point-wise soft shrinkage operator:

$$\mathbf{d}_{k}^{(i+1)} = \text{softshrink}\left(\left[\Phi\left(\mathbf{x}\right)\right]_{k} + \mathbf{b}_{k}^{(i)}, \frac{1}{\mu}\right), \qquad (19)$$

with

s

oftshrink
$$(x, \gamma) = \frac{x}{|x|} \max(|x| - \gamma, 0)$$
. (20)

Here x is one coefficient and γ is the shrinkage threshold. Shrinkage is an extremely fast operation because it requires only a few operations per element of $\mathbf{d}^{(i+1)}$.

When shearlets are used as sparsifying operator, all calculations involving element-wise operations (such as softshrinkage or the Bregman update step) are executed per element per scale per direction. It was empirically found that the soft-shrinkage is best applied using an energy dependent thresholding, by multiplying the threshold value $\frac{1}{\mu}$ by the energy of the subband filter where the threshold is being applied. This energy is calculated by applying the radial filters of the shearlet transformation to a Dirac delta function and calculating the squared ℓ_2 -norm of the resulting shearlet coefficients per scale. This energy will be the highest for the highest resolution scales, and lowest for the low-resolution scale.

The pseudocode of the split-Bregman algorithm for CT reconstruction can be found in the Appendix. Lines 1-9 correspond to the pre-calculation of the energy dependent thresholding in the case of shearlet-based regularization.

III. DATA ACQUISITION

A. Data simulation

Two phantoms were used to measure the performance and properties of the proposed algorithms. The first simulation phantom consists of a 4 cm diameter phantom, containing two gradients. Both gradients are 1 cm wide, ranging from contrast -85% to 100%. This allows us to evaluate the difference of using TV regularization to that of SH regularization for a smooth, non-piecewise constant object.

The second phantom is the clock phantom used by Evans et al. [67], scaled down to small-animal size. It consists of a 4 cm diameter water background, modeled after the attenuation of water at an energy of 60 keV. The phantom contains 8 inserts, each with a different contrast. Each insert is located 1.1 cm from the phantom center, with a diameter of 0.4 cm. This phantom will be used to determine the optimal choice of $\frac{\mu}{\lambda}$ and the effect of contrast magnitude on the noise-resolution tradeoff.

All phantom data was simulated by forward projecting an oversampled phantom using the geometry of the micro-CT scanner used for the real measurements (Sec. III-B). This will allow us to use the same reconstruction settings for all datasets. We simulated fan-beam data over 512 uniformly spaced angles over 2π . The detector consists of 592 elements with a pixel pitch of 0.2 mm. The distance between tube and detector was set to 300.33 mm and the radius of rotation 113.39 mm, resulting in a zoom factor of 3. The detector was offset by

9.69 pixels. After forward projecting the phantom, photon noise was added corresponding to 2×10^5 photons per pixel. To investigate the influence of angular undersampling (fewer-view reconstruction), this sinogram was reduced to only 128 projection views, by uniformly selecting each 4th projection view.

B. Data acquisition

To evaluate the performance on real data and for realistically textured objects, one realistic preclinical phantom was imaged. The plastimouse phantom (Frank Verhaegen¹, Maastro Clinic, the Netherlands) is made by plastinating a mouse. This ensures that no movement will occur while scanning over long time periods. Although the image contrast obtained with this technique is not realistic because all bodily fluids have been replaced by the same plastic, the textures remain. This allows for in depth evaluation of the proposed reconstruction methods.

A Triumph-II CT scanner (TriFoil Imaging, Northridge, CA, USA) was used to conduct the acquisitions. It consists of a 2368×2240 pixel detector with pixel pitch 0.050 mm. A zoom of $3 \times$ was employed to get a field of view (FOV) of 42 mm. The tube voltage was set to 75 kVp, with a tube current of 240 μ A, a detector exposure time of 700 ms and a focal spot size of 50 μ m. The plastimouse was scanned 33 times in the same position, each time using 1024 angles over 2π . As the reconstruction algorithms were only implemented for fan-beam data, only the central detector row was retained from the measured cone-beam data to give fanbeam data without the need for rebinning. This fanbeam dataset was then subsampled to 592 detectorpixels (pitch 0.2 mm), and 512 angles by uniformly selecting each 2nd projection view.

Three datasets were generated from this acquisition data. The first is a reference (quasi-noiseless) sinogram, generated by averaging all 33 fanbeam sinograms. The second is a noisy sinogram, which was made by selecting only one sinogram of the 33 measured ones. The last dataset is fewer-view data, generated by reducing the reference sinogram to 128 views, keeping each 4th sinogram row. Both the reconstructions of the noisy sinogram and the fewer-view sinogram can then be compared to the reference reconstruction obtained from the quasi-noiseless sinogram.

IV. DATA RECONSTRUCTION AND ANALYSIS

All datasets were reconstructed using SIRT, split-Bregman using isotropic TV (SpBR-TV) and split-Bregman using shearlets (SpBR-SH), to a 512^2 voxel grid. A voxel pitch of 0.080 mm was chosen to include the full phantom in the FOV.

The pseudocode to SpBR found in the Appendix was implemented in C++, using the 2D distance driven algorithm [68] as forward and back projector. The number of CG iterations used in the inner loop was always set to 30. We used $\mathbf{C} = \text{diag}\{c_i\}$ with $c_i = e^{-y_i}$, using y_i as an estimator for the mean number of photons detected in detector pixel *i* [51], even when the ground-truth was available from the simulations. In the case

¹Frank.Verhaegen@maastro.nl

of shearlet-based regularization, four subbands were used with 8 analysis directions each. Parameter α was always set to $\frac{1}{2}$. The energy dependent thresholding was implemented by multiplying $\frac{1}{\mu}$ (γ in Eq. (20)) with the energy of the subband where the threshold is being applied (see the last paragraph of Section II-2).

A. Performance analysis

The optimal convergence parameters were determined by holding the cost function constant (i.e. constant λ), after which the iteration at convergence will be determined for different values of convergent rate $\frac{\mu}{\lambda}$. In a second step, the convergence rate is held constant, while determining the iteration at convergence for different cost functions, determined by varying λ . We define convergence as $\|\mathbf{x}^{(\infty)} - \mathbf{x}^{(i)}\|_2^2 < 10^{-4}$. We chose to approximate $\mathbf{x}^{(\infty)}$ by the reconstructed image at iteration 80.

The effect of contrast magnitude on the noise-resolution tradeoff for SIRT, SpBRTV and SpBRSH will be determined by using the method proposed by Evans et al. [67]. Because the proposed algorithms are non-linear, both noise and resolution need to be evaluated locally. For each contrast value, noise is determined by a circular region around the insert. The resolution is determined by the Edge Spread Function (ESF) of the insert edge. A parameterized ESF model is then fitted to this profile [69], after which this analytical model can be differentiated to generate a Line Spread Function (LSF). By using an analytical model for the ESF, the influence of limited sampling and noise on the LSF calculation is reduced. The LSF is then Fourier transformed to generate the Modulation Transfer Function (MTF). Because the MTF is a function and not a single metric, Evans et al. [67] represent the MTF by the value A_{10} :

$$A_{10} = \frac{\int_0^{10\,\mathrm{lp/mm}}\,\mathrm{MTF}\,(f)\,\mathrm{d}f}{10}.$$
 (21)

 A_{10} is normalized to 10, as this is the area under an ideal MTF curve that has amplitude 1.0 for all spatial frequencies. The higher A_{10} , the higher the resolution.

B. Texture analysis

The influence of SpBR-TV and SpBR-SH on texture was measured on the measured plastimouse data with the Gray Level Co-occurrence Matrix (GLCM) [70], calculated on a region of interest (ROI) encompassing the stomach contents (dashed square in Fig. 4). The GLCM allows us to extract second order statistical texture features. For this evaluation study, we compare 4 metrics defined over the GLCM: contrast, correlation, energy and homogeneity. Contrast is a measure of the intensity contrast between a pixel and its neighbor over the whole image, with 0 the contrast value of a constant image. Correlation is a measure of how correlated a pixel is to its neighbor over the whole image. The energy returns the sum of squared elements in the GLCM. A GLCM with a low number of high-value elements will have a higher energy than a GLCM with a high number of low-value elements. A constant image



Fig. 4. Reference image of the three datasets used. Left: gradient phantom. Middle: contrast phantom. Right: measured plastimouse reconstruction. Dashed square: ROI used for texture analysis.

will have an energy of 1. Finally, the homogeneity represents the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

For each reconstructed image, the stomach ROI was cropped and quantized to 32 levels. The GLCM was then calculated for the 0° , 45° , 90° and 135° neighbors using the functionality provided by MATLAB (MATLAB 7.11.0, The Mathworks Inc., Natick, MA). As a final aggregate texture metric, the 4 texture properties were combined to calculate the Euclidean distance to the texture metric calculated on the reference texture. All 4 properties were normalized so to give equal weight to each property inside the distance function.

V. RESULTS

A. Gradient phantom

Figure 5 shows a zoomed in part of the the reconstructions of the gradient phantom using SIRT, SpBR-TV and SpBR-SH, compared to the reference phantom. Both SpBR-TV and SpBR-SH are shown at matched noise (1.5%), while SIRT is shown at convergence (800 iterations). Both methods correctly reconstruct both the spherical gradient as well as the wedge gradient. On the wedge gradient (bottom rows), it can be noticed that TV-based reconstruction does not follow a smooth line along the gradient, and follows the straight edge too fast. This is a sign of slightly lower resolution for the chosen parameter settings. The shearlet-based reconstruction follows the gradient more smoothly, but exhibits oscillating effects around sudden intensity changes. This is also visible in the image itself, where vertical lines are present around the hard edge.

B. Convergence analysis

The convergence of SpBR-TV and SpBR-SH is plotted in Fig. 6. For SpBR-TV, a convergence rate of 3 is optimal, resulting in convergence in only 10 iterations. The convergence becomes worse with a rate lower or higher than this optimum. For SpBR-SH, the optimum is a lower bound on μ/λ . All values larger than $\mu/\lambda = 9$ lead to convergence in 10 iterations.

Based on these results, the optimal μ/λ of 3 was chosen for SpBR-TV reconstruction, and $\mu/\lambda = 10$ for SpBR-SH. We can now investigate the number of iterations needed for different cost functions. Fig. 6b shows that the number of iterations needed can be estimated, as long as the regularization is not too greatly enforced (large λ). For λ values



Fig. 5. Reconstructions of gradient phantom. TV and SH are matched by noise, SIRT shown at convergence. Top rows: zoomed in on spherical gradient. Middle row: profiles through wedge gradient. Red line is reference reconstruction.



Fig. 6. Convergence analysis for (left) fixed cost function and (right) for fixed convergence rate.

smaller than 100 in the case of SpBR-TV, and smaller than 60 in the case of SpBR-SH, more and more iterations are needed to reach the level of convergence needed. Although the optimal parameters can be selected that ensure convergence in than 10 iterations, in any case we will reconstruct all datasets in the next subsections with 30 SpBR iterations. This will ensure convergence regardless of the λ parameter selected. The convergence rate for the following experiments was set to the optimal $\mu/\lambda = 3$ for SpBR-TV and $\mu/\lambda = 10$ for SpBR-SH.

C. Noise-resolution tradeoff

Figures 7 and 8 plot the noise-resolution tradeoff, for different contrast magnitudes. Each data point was gathered by reconstructing with varying λ , with lower λ leading to less noise. The full-view dataset is the contrast phantom with all 512 views, while the fewer-view dataset is the contrast phantom with just 128 views. For higher-noise data points the ESF model fitting becomes difficult, leading to the erratic behavior. However, the general trend of each curve is still visible.

Three observations can be made. First, generally all SpBR-SH curves are positioned above and left of the SpBR-TV curves reconstructed from the same dataset. Thus, the noise is higher for equal resolution when shearlets are used instead of TV. For some noise level, the lowest resolution is always reached by the SpBR-SH reconstruction method on few-view data (Fig. 8). Both methods outperform SIRT reconstruction, especially on fewer-view data. A little better resolution can



Fig. 7. Influence of contrast magnitude on noise-resolution tradeoff for the full-view, noisy dataset. Dashed line: SIRT. Full line: SpBR-TV. Dotted line: SpBR-SH. At equal resolution, noise is higher with SpBR-SH than with SpBR-TV, especially for the $\pm 7\%$ contrast insert.

be found with SpBR-SH compared to SpBR-TV for fewerview data for the $\pm 7\%$ contrast inserts. A zoom on the -7%insert reconstructed from fewer-view data is provided in Fig. 9. At matched resolution of $A_{10} = 0.6$, little noise is seen for TV, while some patchy structures are visible on the insert edge. For SH, some wavy lines are present that will lead to higher noise measurements, though the insert is visually rounder. At maximum resolution (SIRT $A_{10} = 0.91$, SpBR-TV $A_{10} = 0.88$, SpBR-SH $A_{10} = 0.96$) the insert is still visibly rounder for SH, though the wavy lines are even more visible now.

Secondly, the difference between both SpBR methods becomes more and more clear for lower contrast magnitudes. At low contrast ($\pm 7\%$) more noise is measured at equal resolution for SpBR-SH compared to SpBR-TV. This effect is less visible for $\pm 30\%$ contrast, where at lower resolution both methods obtain the same noise level. Only at higher resolution a separation between the two methods occurs.

Finally, comparing the results of the full-view dataset (Fig. 7) to the results obtained from the fewer-view dataset (Fig.



Fig. 8. Influence of contrast magnitude on noise-resolution tradeoff for the fewer-view, noiseless dataset. Dashed line: SIRT. Full line: SpBR-TV. Dotted line: SpBR-SH. At equal resolution, noise is higher with SpBR-SH than with SpBR-TV, especially for the $\pm 7\%$ contrast insert.



Fig. 9. -7% contrast insert from the few-view dataset reconstructed with SIRT, SpBR-TV and SpBR-SH at matched resolution ($A_{10} = 0.6$) and at maximum resolution. Window [0.18 0.23] cm⁻¹.

8), makes it clear that a higher resolution can be achieved at the same noise level for full-view datasets. When fewer-view data is used, the resolution drops. This is true for all three the methods.

D. Texture analysis

Figure 10 plots the 4 texture metrics which were determined from the GLCM. Each plot represents one texture metric, and contains the results for SpBR-TV (full line) and SpBR-SH



Fig. 10. Contrast, correlation, energy and homogeneity as measured from the GLCM. Full line: SpBR-TV. Dotted line: SpBR-SH. Dashed line: reference value obtained from full-view noiseless SIRT reconstruction. Black: noisy dataset. Gray: fewer-view dataset.

(dotted line) for both the noisy dataset (black) as well as the fewer-view dataset (gray). The value obtained for the averaged plastimouse dataset with all views is pictured by the dashed line, which serves as a reference property value. The bottom row contains two plots that represent the Euclidean distance of the 4 selected properties to the reference reconstruction. Different data point were gathered by reconstructing with varying λ in the case of SpBR-methods, where for SIRT this is done by varying the number of iterations. The noise is measured in the ROI depicted by the dashed ellipse in Fig. 11.

Here, shearlet-regularized reconstruction outperforms TVbased reconstruction, for both the noisy as well as the fewerview dataset. TV is always closer to the reference than SH on the contrast metric. Based on correlation, both methods perform similarly independent of the dataset, though the same correlation value is reached at higher noise for SpBR-SH compared to SpBR-TV. A big difference can be found for the energy property. Here, while SH estimates the energy correctly, TV regularization removes the structure completely and replaces it by homogeneous patches. This results in a very high energy, which results in a high distance to the reference texture. The same is visible to a lesser extent on the homogeneity property.

Comparing both distance plots in the lowest row of Fig. 10, the distance for the noisy dataset has a clear optimum for the shearlet reconstruction. For noisy data, the texture obtained



Fig. 11. Reconstructions for the noisy dataset. Reconstruction images matched on noise, measured in a ROI (dashed ellipse) on the liver.



Fig. 12. Reconstructions for the fewer-view dataset. Reconstruction images matched on noise, measured in a ROI (dashed ellipse) on the liver.

VI. DISCUSSION

with shearlet-based reconstruction is closer to the reference than obtained with conventional SIRT reconstruction, at lower noise. With fewer-view data, SpBR-SH outperforms SIRT. When more regularization is applied, the distance to the reference reconstruction increases, and this leads to a worse texture than when normal SIRT is used. However, these lownoise SIRT reconstructions are generated with only a few iterations, which means the corresponding resolution will be far worse than the converged SpBR-SH reconstructions.

Figure 11 shows reconstructions of the plastimouse for the noisy dataset, comparing to the reference SIRT reconstruction generated from the 33× averaged sinogram. The SIRT, SpBR-TV and SpBR-SH reconstructions were selected with an equal noise level (12%) by experimentally selecting the correct λ value which gives this noise level at convergence for SpBR-methods, and the early-stopping of SIRT when this noise level was reached. Only SpBR-SH can accurately reconstruct the small diagonal stripes seen in the left kidney (full arrow). SpBR-TV exhibits patchiness which is especially visible on object edges, and in the darker region on the right of the image (dashed arrow). Edges are more smoothly reconstructed when SpBR-SH is used. However, the darker region shows some wavy artifacts when shearlets are used, compared to the blocky structures encountered in the TV-based reconstruction.

The fewer-view dataset reconstructions are shown in Figure 12, matched to 10% noise in the liver ROI. Both SpBR-TV and SpBR-SH have problems with reducing the noise in this dataset while keeping the texture of the stomach intact. Both methods still show a lot of aliasing artifacts, caused by the 4-fold undersampling of the data. For SpBR-SH, the aliasing streaks have been sharpened in some places, while they have been minimized in others. Small spots (full arrow) are better visible on the TV-regularized reconstruction than with SpBR-SH, where it has been smoothed away.

Replacing TV by shearlets for regularized reconstruction is promising for iterative CT imaging in theory. We have shown that shearlets do not lead to the piecewise-constant behavior as seen with TV, but instead may lead to wavy artifacts, which can be equally unwanted in medical diagnostics.

The noise-resolution tradeoff shows that SpBR-SH leads to higher noise at the same resolution as obtained with SpBR-TV. However, this could be due to the Gibbs effect also noticed in Fig. 5. These wavy oscillations will be included in the noise measurements, leading to a bias in the tradeoff curves. These wavy lines are also visible on Fig. 9 and on the real data presented in Fig. 11 and Fig. 12. They replace the patchy artefacts of TV regularization.

Unfortunately, it is very difficult to quantitatively evaluate the added usefulness from the medical diagnostic point of view, as a ground-truth reference image will always be needed. The most-used phantom in CT reconstruction evaluation is the Shepp-Logan phantom, defined by adding 10 uniformly-filled ellipses [13]. This makes it inherently piecewise constant. Promising is the work done by Bond et al. [16], who have developed a version of the XCAT phantom [27] with realistic lung texture. However, it still needs to be expanded into 3D texture. In our study presented here, we used a high-resolution, realistically textured, plastinated mouse, so that the dataset would not be biased towards TV.

Although the texture analysis on preclinical images improved the outlook for shearlet-based regularization, it is clear that there are still some issues. Although the shearlet transform, like the curvelet transform, is suited for structures representing sharp and elongated structures, e.g. edges, it is unsuited for spherical-like sources (the point in the stomach on Fig. 12 has been reconstructed less accurately in SpBR-SH). This is a common issue with a lot of multi-resolution techniques, as when they represent isotropic features well, they are far from optimal for analyzing anisotropic objects [71]. This was the main idea behind the development of other constructions, such as the curvelet transform [72]. It could thus be beneficial to combine the shearlet transformation with an isotropic wavelet transform. The sensitivity of each transform to a particular shape would possibly make it a very strong discriminating tool. However, the weights of the transformations in this combination will depend on the specific study.

Although the methods presented here have been applied to fan-beam data, the techniques can be carried over to 3D conebeam data with no further adjustments. Indeed, no assumptions have been made about the scanner geometry in the derivation of the algorithms, and the optimization algorithms used can be applied irrespectively from dimensionality. Although the shearlet methodology can be extended to 3D directly [73], it will suffer from large computational complexity due to the use of multiple 3D FFTs and subsequent frequency domain filtering (which has to be repeated for every iteration of the split-Bregman algorithm). A possible alternative is a special 3D shearlet design based on separable filters in the spatial domain. This is a topic of ongoing research. Even then, there is still room to improve the existing algorithm, by using preconditioned CG [74]. This would improve the reconstruction time by reducing the number of CG iterations needed.

Future work could include the application of these developed techniques on different datasets together with numerical observer studies, e.g. the channelized Hotelling observer [75], [76], to compare different regularization strategies from the medical diagnostic point of view.

VII. CONCLUSION

We have presented a split-Bregman based algorithm to perform iterative CT reconstruction, using shearlet regularization as an alternative for TV minimization. The use of shearlets for regularization leads to different artifacts than in the case of TV, because shearlets model the structures contained in the image using a different (non-piecewise constant) image model. Based on a noise-resolution tradeoff study, TV minimization outperforms shearlet-based reconstruction. However, on acquired data with realistic textures, shearlets reconstruct textures more similar to the reference texture than when TV is used. The piecewise constant artifacts are gone, but have been replaced by wave-like structures.

ACKNOWLEDGEMENT

This work was supported in part by a PhD grant to Bert Vandeghinste of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) and by the CIMI project, an iMinds-project. iMinds is an independent multidisciplinary research institute founded by the Flemish Government, to stimulate ICT innovation. Bart Goossens and Roel Van Holen are postdoctoral research fellows of the Research Foundation Flanders (FWO). Christian Vanhove is supported by the GROUP-ID consortium. The authors would also like to thank Gabor T. Herman and Jason Bond for providing us the phantoms used in Fig. 1.

11

APPENDIX

The following algorithm shows the pseudocode of the split-Bregman algorithm for regularized CT reconstruction.

Require: $\mathbf{d} \leftarrow \mathbf{0}, \mathbf{b} \leftarrow \mathbf{0}, \mathbf{C} \leftarrow \text{diag}(e^{-y_i}), \mu, \lambda$. Ensure: image x. 1: if Φ = Shearlet-transform then $\mathbf{u} = \mathbf{\Phi}(\mathbf{1})$ 2: for $s = 0 \rightarrow scales(\mathbf{u})$ do 3: for $i = 0 \rightarrow size(\mathbf{u}[s])$ do $\mathbf{E}[s][i] = \sqrt{\sum_{j=0}^{size(\mathbf{u}[s])} \mathbf{u}[s][j]^2}$ end for 4: 5: 6: 7: end for 8: else 9: $\mathbf{E} = \mathbf{1}$ 10: end if 11: for $0 \rightarrow iterations$ do $\mathbf{r} \leftarrow \mathbf{W}^{\dagger}(\mathbf{C}^{-1}\mathbf{y}) + (\mu/\lambda)\mathbf{\Phi}(\mathbf{d} - \mathbf{b})$ 12: $\mathbf{p} \leftarrow \mathbf{r}$ 13: $new_err \gets \mathbf{r}' * \mathbf{r}$ 14: for $0 \rightarrow cg_iterations$ do 15: $err \leftarrow new \ err$ 16: $\mathbf{M} \leftarrow \mathbf{W}^{\dagger}(\mathbf{C}^{-1}\mathbf{W}(\mathbf{p})) + (\mu/\lambda)\mathbf{\Phi}^{\dagger}(\mathbf{\Phi}(\mathbf{p}))$ 17: $\alpha \leftarrow err/(\mathbf{p}' * \mathbf{M})$ 18: $\mathbf{x} \leftarrow \mathbf{x} + \alpha \mathbf{p}$ 19: $\mathbf{r} \leftarrow \mathbf{r} - \alpha \mathbf{M}$ 20: *new* $err \leftarrow \mathbf{r}' * \mathbf{r}$ 21: $\beta \leftarrow new_err/err$ 22: $\mathbf{p} \leftarrow \mathbf{r} + \beta \mathbf{p}$ 23: end for 24: $\mathbf{S} = \mathbf{\Phi}(\mathbf{x})$ 25: $\mathbf{k} = \mathbf{S} + \mathbf{b}$ 26: for $i = 0 \rightarrow size(\mathbf{k})$ do 27: $\mathbf{d}[i] \leftarrow \operatorname{sign}(\mathbf{k}[i]) * \max(|\mathbf{k}[i]| - \mathbf{E}[i]/\mu, 0)$ 28: end for 29. $\mathbf{b} \leftarrow \mathbf{b} + (\mathbf{S} - \mathbf{d})$ 30: 31: end for 32: return x

REFERENCES

- L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.
- [2] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *Journal* of X-Ray Science and Technology, vol. 14, no. 2, pp. 119–139, 2006.
- [3] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization." *Physics in Medicine and Biology*, vol. 53, no. 17, pp. 4777–4807, Sep. 2008.
- [4] E. Sidky, X. Pan, I. Reiser, R. Nishikawa, R. Moore, and D. Kopans, "Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms," *Medical Physics*, vol. 36, p. 4920, 2009.
- [5] G. T. Herman and R. Davidi, "On Image Reconstruction from a Small Number of Projections." *Inverse Problems*, vol. 24, no. 4, pp. 45011– 45028, Aug. 2008.
- [6] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical Physics*, vol. 35, no. 2, pp. 660–663, 2008.
- [7] H. Yu and G. Wang, "Compressed sensing based interior tomography." *Physics in Medicine and Biology*, vol. 54, no. 9, pp. 2791–2805, May 2009.

- [8] J. Yang, H. Yu, M. Jiang, and G. Wang, "High-order total variation minimization for interior tomography," *Inverse Problems*, vol. 26, no. 3, p. 035013, Feb. 2010.
- [9] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Adaptative Total Variation Image Deblurring: A Majorization-Minimization Approach," *Signal Process*, vol. 89, pp. 1683–1693, 2009.
- [10] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.
- [11] Y. Meyer, Oscillating patterns in image processing and nonlinear evolution equations, ser. Univ. Lecture Ser. 22. American Mathematical Soc., 2001.
- [12] W. Yin, D. Goldfarb, and S. Osher, "Total variation based image cartoontexture decomposition," *SIAM Journal on Multiscale Modeling and Simulation*, 2005.
- [13] L. A. Shepp and B. F. Logan, "Reconstructing Interior Head Tissue from X-Ray Transmissions," *IEEE Transactions on Nuclear Science*, vol. 21, no. 1, pp. 228–236, 1974.
- [14] E. J. Candes, J. K. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Ieee Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [15] E. Garduño, G. T. Herman, and R. Davidi, "Reconstruction from a Few Projections by L1-Minimization of the Haar Transform." *Inverse Problems*, vol. 27, no. 5, p. 055006, May 2011.
- [16] J. Bond, D. Frush, E. Samei, and W. P. Segars, "Simulation of anatomical texture in voxelized XCAT phantoms," *SPIE Medical Imaging Sympo*sium, 2013.
- [17] J. Tang, B. E. Nett, and G.-H. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms," *Physics in Medicine and Biology*, vol. 54, no. 19, pp. 5781–5804, 2009.
- [18] J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT," *Medical Physics*, vol. 34, no. 11, pp. 4476–4483, 2007.
- [19] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [20] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magnetic Resonance in Medicine*, vol. 61, no. 1, pp. 103–116, Jan. 2009.
- [21] A. S. Carasso, "Singular integrals, image smoothness, and the recovery of texture in image deblurring," *SIAM Journal on Applied Mathematics*, pp. 1749–1774, 2004.
- [22] Y. Gousseau and J.-M. Morel, "Are natural images of bounded variation?" Siam Journal on Mathematical Analysis, vol. 33, no. 3, pp. 634– 648, 2001.
- [23] A. Buades, B. Coll, and J.-M. Morel, "The staircasing effect in neighborhood filters and its solution," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1499–1505, 2006.
- [24] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazon, "Minimizing total variation flow," *Differential and Integral Equations*, vol. 14, no. 3, pp. 321–360, 2001.
- [25] F. Andreu, V. Caselles, J. I. Diaz, and J. M. Mazon, "Some qualitative properties for the total variation flow," *Journal of Functional Analysis*, vol. 188, no. 2, pp. 516–547, 2002.
- [26] A. Chambolle and P. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, vol. 76, no. 2, pp. 167–188, 1997.
- [27] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Medical Physics*, vol. 37, no. 9, p. 4902, 2010.
- [28] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs," *SIAM Journal on Numerical Analysis*, vol. 42, no. 2, pp. 686–713, 2005.
- [29] I. Selesnick, R. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *Signal Processing Magazine*, *IEEE*, vol. 22, no. 6, pp. 123–151, 2005.
- [30] E. J. Candes, "Ridgelets: theory and applications," Ph.D. dissertation, Stanford University, 1998.
- [31] M. N. Do and M. Vetterli, "The finite ridgelet transform for image representation." *IEEE Transactions on Image Processing*, vol. 12, no. 1, pp. 16–28, 2003.

- [32] E. J. Candes, L. Demanet, D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling and Simulation*, vol. 5, no. 3, pp. 861–899, 2006.
- [33] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising." *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.
- [34] K. Guo and D. Labate, "Optimally sparse multidimensional representation using shearlets," *Siam Journal on Mathematical Analysis*, vol. 39, no. 1, pp. 298–318, 2007.
- [35] K. Guo, D. Labate, and W.-Q. Lim, "Edge analysis and identification using the continuous shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 27, no. 1, pp. 24–46, 2009.
- [36] F. Bochud, C. K. Abbey, and M. Eckstein, "Statistical texture synthesis of mammographic images with clustered lumpy backgrounds," *Optics Express*, vol. 4, no. 1, pp. 33–43, 1999.
- [37] B. Goossens, "Multiresolution image models and estimation techniques," Ph.D. dissertation, Ghent University, Ghent, 2010.
- [38] G. R. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, 2008.
- [39] G. R. Easley, D. Labate, and F. Colonna, "Shearlet-Based Total Variation Diffusion for Denoising," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 260–268, 2009.
- [40] E. Tadmor, S. Nezzar, and L. Vese, "A Multiscale Image Representation Using Hierarchical (BV, L2) Decompositions," *Multiscale Modeling and Simulation*, vol. 2, no. 4, pp. 554–579, Jan. 2004.
- [41] E. J. Candes and D. L. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," *Annals of statistics*, pp. 784– 842, 2002.
- [42] B. Feng, S. Kunis, and D. Potts, "On the computation of the polar FFT," *Applied and Computational Harmonic Analysis*, vol. 22, no. 2, pp. 257–263, 2007.
- [43] J. Aelterman, H. Q. Luong, B. Goossens, A. Pižurica, and W. Philips, "Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data," *Signal Processing*, vol. 91, no. 12, pp. 2731–2742, Dec. 2011.
- [44] B. Vandeghinste, B. Goossens, R. Van Holen, C. Vanhove, A. Pižurica, S. Vandenberghe, and S. G. Staelens, "Iterative CT reconstruction using shearlet-based regularization," *Proc. of SPIE*, pp. 1–7, 2012.
- [45] F. Colonna, G. R. Easley, K. Guo, and D. Labate, "Radon transform inversion using the shearlet representation," *Applied and Computational Harmonic Analysis*, vol. 29, no. 2, pp. 232–250, Sep. 2010.
- [46] X.-C. Tai and C. Wu, "Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model," *Scale Space and Variational Methods in Computer Vision*, pp. 502–513, 2009.
- [47] T. Goldstein and S. Osher, "The split Bregman method for L1 regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323– 343, 2009.
- [48] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *Journal of Theoretical Biology*, vol. 36, no. 1, pp. 105–117, Jul. 1972.
- [49] G. T. Herman and A. Lent, "Iterative reconstruction algorithms," Computers in Biology and Medicine, vol. 6, no. 4, pp. 273–294, Oct. 1976.
- [50] A. Wunderlich and F. Noo, "Image covariance and lesion detectability in direct fan-beam x-ray computed tomography." *Physics in Medicine and Biology*, vol. 53, no. 10, pp. 2471–2493, May 2008.
- [51] K. D. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *Signal Processing, IEEE Transactions* on, vol. 41, no. 2, pp. 534–548, 1993.
- [52] J. Hsieh, Computed Tomography: Principles, Design, Artifacts, and Recent Advances, 2nd ed. SPIE, 2009.
- [53] J. Wang, A. Chai, and L. Xing, "Noise correlation in CBCT projection data and its application for noise reduction in low-dose CBCT," *Proceedings of the SPIE*, vol. 7258, pp. 72582D1–8, 2009.
- [54] H. Yu and G. Wang, "A soft-threshold filtering approach for reconstruction from a limited number of projections," *Physics in Medicine and Biology*, vol. 55, pp. 3905–3916, 2010.
- [55] R. D. Nowak and M. A. T. Figueiredo, "Fast wavelet-based image deconvolution using the EM algorithm," in *Proceedings of the 35th Asilomar Conference on Signals, Systems, and Computers.* Monterey, CA: IEEE, Nov. 2001, pp. 371–375.
- [56] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

- [57] A. Pižurica, J. Aelterman, F. Bai, S. Vanloocke, H. Quang Luong, B. Goossens, and W. Philips, "On structured sparsity and selected applications in tomographic imaging," *Proc. of SPIE Conference Wavelets* and Sparsity XIV, p. 81381D, 2011.
- [58] Y. Wang, W. Yin, and Y. Zhang, "A fast algorithm for image deblurring with total variation regularization," *CAAM Technical Reports*, 2007.
- [59] D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss, "Sparse multidimensional representation using shearlets," *Proceedings of the SPIE*, 2005.
- [60] G. Kutyniok and D. Labate, "Resolution of the Wavefront Set Using Continuous Shearlets," *Transactions of the American Mathematical Society*, vol. 361, no. 5, pp. 2719–2754, 2009.
- [61] S. Yi, D. Labate, G. R. Easley, and H. Krim, "A Shearlet Approach to Edge Analysis and Detection," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 929–941, 2009.
- [62] B. Goossens, J. Aelterman, H. Q. Luong, A. Pižurica, and W. Philips, "Efficient design of a low redundant discrete shearlet transform," *Proc.* of the 2009 Int. Workshop on Local and Non-Local approximation in Image Processing, pp. 112–124, 2009.
- [63] —, "Design of a Tight Frame of 2D Shearlets Based on a Fast Non-iterative Analysis and Synthesis Algorithm," in *IEEE Optics & Photonics 2011, Wavelets and Sparsity XIV*, San Diego, CA, USA, 2011, p. 81381Q.
- [64] ——, "An Efficient Design of a Discrete Shearlet Transform with Bandlimited Basis Functions," *in preparation*.
- [65] I. Daubechies, Ten lectures on wavelets . SIAM, 1992.
- [66] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, no. 6, pp. 409–436, 1952.
- [67] J. D. Evans, D. G. Politte, B. R. Whiting, J. A. O'Sullivan, and J. F. Williamson, "Noise-resolution tradeoffs in x-ray CT imaging: A comparison of penalized alternating minimization and filtered backprojection algorithms," *Medical Physics*, vol. 38, no. 3, p. 1444, 2011.
- [68] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions." *Physics in Medicine and Biology*, vol. 49, no. 11, pp. 2463–2475, 2004.
- [69] J. M. Boone and J. A. Seibert, "An analytical edge spread function model for computer fitting and subsequent calculation of the LSF and MTF." *Medical Physics*, vol. 21, no. 10, pp. 1541–1545, Oct. 1994.
- [70] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 6, pp. 610–621, 1973.
- [71] J.-L. Starck and J. Bobin, "Astronomical Data Analysis and Sparsity: From Wavelets to Compressed Sensing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1021–1030, 2009.
- [72] E. J. Candes and D. L. Donoho, "Ridgelets: A key to higher-dimensional intermittency?" *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 357, no. 1760, pp. 2495–2509, 1999.
- [73] B. Goossens, H. Q. Luong, J. Aelterman, A. Pižurica, and W. Philips, "Efficient multiscale and multidirectional representation of 3D data using the 3D discrete shearlet transform," in *SPIE Optics & Photonics*. International Society for Optics and Photonics, 2011, pp. 81 381Z–13.
- [74] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction." *IEEE Transactions* on *Image Processing*, vol. 8, no. 5, pp. 688–699, 1999.
- [75] K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model." *Journal of the Optical Society of America. A, Optics and image science*, vol. 4, no. 12, pp. 2447–2457, Dec. 1987.
- [76] L. Platiša, B. Goossens, E. Vansteenkiste, S. Park, B. D. Gallas, A. Badano, and W. Philips, "Channelized Hotelling observers for the assessment of volumetric imaging data sets." *Journal of the Optical Society of America. A, Optics, image science, and vision*, vol. 28, no. 6, pp. 1145–1163, Jun. 2011.