Adaptive Graph-Based Total Variation for Tomographic Reconstructions

Faisal Mahmood, Nauman Shahid, Ulf Skoglund*, and Pierre Vandergheynst*

***Abstract*—Sparsity exploiting image reconstruction (SER) methods have been extensively used with total variation (TV) regularization for tomographic reconstructions. Local TV methods fail to preserve texture details and often create additional artifacts due to over-smoothing. Nonlocal TV (NLTV) methods have been proposed as a solution to this but they either lack continuous updates due to computational constraints or limit the locality to a small region. In this letter, we propose adaptive graph-based TV. The proposed method goes beyond spatial similarity between different regions of an image being reconstructed by establishing a connection between similar regions in the entire image regardless of spatial distance. As compared to NLTV, the proposed method is computationally efficient and involves updating the graph prior during every iteration making the connection between similar regions stronger. Moreover, it promotes sparsity in the wavelet and graph gradient domains. Since TV is a special case of graph TV, the proposed method can also be seen as a generalization of SER and TV methods.**

***Index Terms*—Graphs, iterative image reconstruction, nonlocal total variation, tomography, total variation.**

I. Introduction

RECONSTRUCTING tomographic densities from low-dose electron tomography (ET) or computed tomography (CT) data is an *ill-posed inverse problem*. Low-dose is a constraint to prevent sample degradation in ET [1], [2] and to reduce exposure to ionizing radiation in CT [3]–[5]. Such requirements are often met by collecting limited or low-contrast data which renders noisy and erroneous reconstructions. Iterative image reconstruction (IIR) methods [6]–[10] have proved to be more effective in handling noise when compared to analytical methods [11]–[13]. However, such methods are computationally inefficient. Initial IIR methods were algebraic in nature [14]–[19]. More recently, sparsity exploiting reconstructions have been extensively used for image reconstruction. Such methods are often used with total variation (TV) regularization [20]–[25]. We refer to the joint compressed sensing (CS) and TV setup as CSTV in the sequel. Recently, nonlocal TV (NLTV) [26] has been shown to be much more efficient for *inverse problems* [27]–[31]. In contrast to simple TV, which takes into account the similarity of a region with only its neighboring regions, NLTV overcomes this limitation by associating a similarity measure of every region of an image with all other regions (full NLTV) or a few regions in a spatial neighborhood (partial NLTV).

A primary shortcoming of full NLTV is the high cost of associating a similarity measure between every pair of regions in an image ($\mathcal{O}(n^4)$ for an $n \times n$ image). Hence, the similarity matrix constructed in the beginning from the initial estimate or *prior* is not updated throughout the algorithm [29], [30]. In order to overcome the computational complexity for adaptive updates, partial NLTV methods [26], [27] tend to limit the nearest neighbors search to a local neighborhood of the pixel (hence we call them partial), which depends on a parameter δ. For such methods, the computational cost drops down from $\mathcal{O}(n^4)$ to $\mathcal{O}(n^2\delta^2)$, where $\delta \ll n$. However, it is quite probable that two spatially distant patches in an image are quite similar in structure. Thus, such methods lack the capability to model the pairwise relationships between the patches of an image on a global level. The final reconstruction would be more faithful to the data if 1) the similarity matrix is regularly updated during every iteration and 2) pairwise relationships are taken into account among all the patches of the image.

Introduction to graphs: Graphs, a discrete way of characterizing nonlocal variation methods, have emerged as a very powerful tool for signal modeling [32], [33]. A graph is represented as a tuple $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{W}\}$, where $\mathcal{V}$ is a set of vertices, $\mathcal{E}$ a set of edges, and $\mathcal{W} : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}_+$ a weight function. The weight matrix W is assumed to be nonnegative, symmetric, and with a zero diagonal. Each entry of the weight matrix $W \in \mathbb{R}_+^{|\mathcal{V}|\times|\mathcal{V}|}$ corresponds to the weight of the edge connecting the corresponding vertices: $W_{i,j} = \mathcal{W}(v_i, v_j)$, and if there is no edge between two vertices, the weight is set to 0. For a vertex $v_i \in \mathcal{V}$, the degree $d(i)$ is defined as the sum of the weights of incident edges: $d(i) = \sum_{j \leftrightarrow i} W_{i,j}$. Let D be the diagonal degree matrix with diagonal entries $D_{ii} = d(i)$, then the graph Lapla-

Manuscript received December 27, 2017; revised February 22, 2018; accepted March 6, 2018. Date of publication March 16, 2018; date of current version April 16, 2018. The work of F. Mahmood was supported in part by the Japanese Government OIST Subsidy for Operations (Skoglund U.) under Grant 5020S7010020 and in part by the OIST PhD Fellowship. The work of N. Shahid and P. Vandergheynst was supported by the Swiss National Science Foundation under Grant 200021_154350/1. The work of U. Skoglund was supported by the Japanese Government OIST Subsidy for Operations (Skoglund U.) under Grant 5020S7010020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mehdi Moradi. *(Faisal Mahmood and Nauman Shahid contributed equally to this work.) (Ulf Skoglund and Pierre Vandergheynst are equal cosenior authors.) (Corresponding author: Faisal Mahmood.)*

F. Mahmood is with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: faisalm@jhu.edu).

U. Skoglund is with the Structural Cellular Biology Unit, Okinawa Institute of Science and Technology, Onna 904-0495, Japan (e-mail: ulf.skoglund@oist.jp).

N. Shahid and P. Vandergheynst are with the Signal Processing Laboratory 2 (LTS2), École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland (e-mail: nauman.shahid@epfl.ch; pierre.vandergheynst@epfl.ch).

This letter has supplementary downloadable material available at http://ieeexplore.ieee.org.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2018.2816582

cian L is defined as the difference of the weight matrix W from the degree matrix D, thus $L = D - W$, which is referred to as combinatorial Laplacian. A more detailed account of the theory of signal processing on graphs can be found in seminal letters [32]–[36].

Contributions: Our previous work [37] has focused on using graph-based TV for denoising the sinogram as a preprocessing step followed by using standard reconstruction methods such as Simultaneous Iterative Reconstruction Technique (SIRT) or Algebraic Reconstruction Techniques (ART) for reconstruction. In this letter, we propose Adaptive Graph Total Variation (AGTV) as a novel method for simultaneous reconstruction and denoising of tomographic data. Our proposed method can be seen as a more sophisticated and adaptive form of full NLTV in the sense that it enjoys a relatively lower computational complexity by using an approximate $\mathcal{K}$-nearest neighbor search algorithm, where $\mathcal{K}$ is fixed. Due to a significant computational cost reduction, we can afford to update the graph in every iteration making the setup adaptive. Furthermore, our proposed method models the sparsity of the reconstructed image in: 1) wavelet domain and 2) graph gradient domain. These improvements lead to state-of-the art reconstruction results for both the phantom data with known ground truth and the real ET data.

II. Adaptive Graph Total Variation (AGTV)

Let $S \in \Re^{p\times q}$ be the sinogram corresponding to the projections of the sample $X \in \Re^{n\times n}$ being imaged, where p is the number of rays passing through X and q is the number of angular variations at which X has been imaged. Let $b \in \Re^{pq}$ be the vectorized measurements or projections ($b = \text{vec}(S)$), where $\text{vec}(\cdot)$ denotes the vectorization operation and $A \in \Re^{pq\times n^2}$ be the sparse projection operator. Then, the goal in a typical CT- or ET-based reconstruction method is to recover the vectorized sample $x = \text{vec}(X)$ from the projections b. We propose

$$\min_x \|Ax - b\|_2^2 + \lambda\|\Phi^*(x)\|_1 + \gamma\|\nabla_{\mathcal{G}}(x)\|_1 \qquad (1)$$

where Φ is the wavelet operator and $\Phi^*(x)$, where $*$ represents the adjoint operation, denotes the wavelet transform of x, and $\|\nabla_{\mathcal{G}}(x)\|_1$ denotes the TV of x w.r.t graph $\mathcal{G}$. The first two terms of the objective function above comprise the *sparse reconstruction* part of our method and model the sparsity of the wavelet coefficients. The second term, to which we refer as the *graph total variation* (GTV) regularizer acts as an additional prior for denoising and smoothing. It can be expanded as

$$\|\nabla_{\mathcal{G}}(x)\|_1 = \sum_i \|\nabla_{\mathcal{G}} x_i\|_1 = \sum_i \sum_{j\in\mathcal{N}_i} \sqrt{W_{ij}}\|x_i - x_j\|_1$$

where the second sum runs over all the neighbors of i, denoted by $\mathcal{N}_i$. The above expression states that GTV involves the minimization of the sum of the gradients of the signals on the nodes of the graphs. In our case, we assume that the elements of the vector x lie on the nodes of the graph $\mathcal{G}$ which are connected with the edges whose weights are W_{ij}. Thus, the minimization of the GTV would ensure that x_i and x_j possess similar values if W_{ij} is high and dissimilar values if W_{ij} is small or zero. As compared to standard TV, the structure of the sample x is taken into account for reconstruction. It is a well-known fact that l_1 norm promotes sparsity, so the GTV can also be viewed as a regularization which promotes sparse graph gradients. This corresponds to enforcing a piecewise smoothness of the signal x w.r.t graph $\mathcal{G}$.

The proposed method with GTV can be seen as a generalization of the CS- and TV-based method studied in [24]. While, the standard TV minimizes the gradients of the signal x w.r.t its spatial neighbors only, the GTV does so in a region which is not restricted only to the neighbors of the elements in x. Thus, the standard TV can be viewed as a specific case of the GTV, where the graph $\mathcal{G}_{\text{grid}}$ is a grid graph. In a grid graph $\mathcal{G}_{\text{grid}}$ of a sample x, the pixels are only connected to its spatial neighbors via unity weights.

An important step for our method is to construct a graph $\mathcal{G}$ for GTV regularization. Ideally, $\mathcal{G}$ should be representative of the reconstructed sample x; however, this is unknown before the reconstruction. To cater this problem, we propose to construct $\mathcal{G}$ from the patches of an initial naive estimate of the sample x_{fbp} using analytical filtered back projection (FBP). In the first step $x_{\text{fbp}} \in \mathbb{R}^{n\times n}$ is divided into n^2 overlapping patches. Let s_i be the patch of size $l \times l$ centered at the ith pixel of x_{fbp} and assume that all patches are vectorized, i.e, $s_i \in \Re^{l^2}$. In the second step, the search for the closest neighbors for all vectorized patches is performed using the Euclidean distance metric. For two patches s_i, s_j, the distance metric is defined as $\|s_i - s_j\|_2$. Each s_i is connected to its $\mathcal{K}$-nearest neighbors s_j only, resulting in $|\mathcal{E}|$ number of connections. This is realized by computing all the pairwise distances between all possible patches s_i, s_j and then keeping only the most relevant $\mathcal{K}$ neighbors. In the third step, the graph weight matrix W is computed using the Gaussian kernel weighting scheme, for which the parameter σ is set experimentally as the average distance of the connected samples. Hence, for the patches s_i, s_j, the weighting scheme is defined as $W_{i,j} = \exp(-\|s_i - s_j\|_2^2/\sigma^2)$. Finally, the combinatorial Laplacian is computed.

Note that the computation of the weight matrix W for graph $\mathcal{G}$ costs $\mathcal{O}(n^4)$. As mentioned earlier, our goal is to avoid this cost and update the graph in every iteration. For this purpose, we propose to make the graph construction efficient by using an approximate nearest-neighbor search algorithm by using the FLANN library (Fast Library for Approximate Nearest Neighbors searches in high-dimensional spaces) [38]. This reduces the cost of graph construction from $\mathcal{O}(n^4)$ to $\mathcal{O}(n^2\log(n))$.

The above description refers only to the nonadaptive part, where the graph $\mathcal{G}$ is fixed. It is important to point out that the initial estimate of the graph $\mathcal{G}$, obtained via the FBP x_{fbp} is not very faithful to the final solution x. As x is being refined in every iteration, it is natural to update the graph $\mathcal{G}$ as well in every iteration. This simultaneous update of the graph $\mathcal{G}$ corresponds to the adaptive part of the proposed algorithm and its significance has been explained in detail in the supplement with this letter.

III. Optimization Solution

In the spirit of similar nongraph methods such as [24], we refer to (1) without the graph update as CS and GTV or simply GTV. We make use of forward–backward-based primal dual

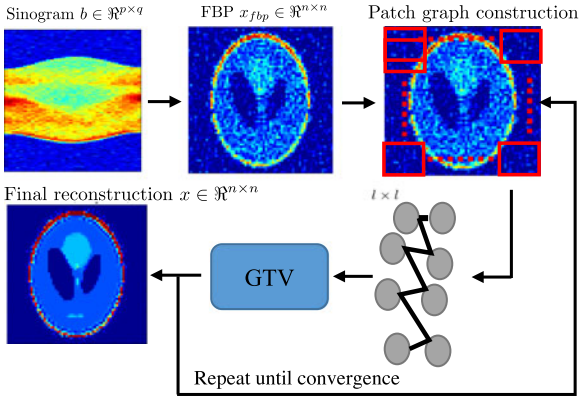

Fig. 1. Complete methodology for AGTV. The input projections $b \in \mathbb{R}^{p\times q}$ is first used to obtained a filtered back projection (FBP) $x_{\text{fbp}} \in \mathbb{R}^{n\times n}$. It is then used to construct the initial patch graph $\mathcal{G}$ to be used by the GTV method. The output of GTV is used to refine/reconstruct the graph and this process is repeated until convergence.

method [39], [40] to solve GTV and then update the graph from the obtained sample in every iteration, until convergence. The complete algorithm with graph updates is called AGTV. The main steps of this algorithm are visualized in Fig. 1.

The first term of (1), $f : \mathbb{R}^{n^2} \to \mathbb{R}$ is a convex differentiable function defined as $f(x) = \|Ax - b\|_2^2$. This function has a β-Lipschitz continuous gradient $\nabla_f(x) = 2A^\top(Ax - b)$. Note that $\beta = 2\|A\|_2$ where $\|A\|_2$ is the spectral norm (or maximum eigenvalue) of A. The constant β has important implications in deciding the time step in iterative optimization methods. Let τ_1, τ_2, τ_3 be the step size parameters. As a rule of thumb, these parameters are typically set to the inverse of the Lipschitz constant β. Hence, we set τ_1, τ_2, τ_3 proportional to $1/\beta$. Furthermore, note that these parameters are independent of the regularization parameters λ and γ.

The proximal operator of the second function $g = \lambda\|\Phi^*(x)\|_1$ [in (1)] is the ℓ_1 soft thresholding of the wavelet coefficients given by the elementwise operations.

$$\text{prox}_{\tau_1 g}(x) = \text{sgn}(x) \circ \max(|x| - \tau_1\lambda, 0). \tag{2}$$

The third term in (1) $h : \mathbb{R}^{|\mathcal{E}|n} \to \mathbb{R}$, where $|\mathcal{E}|$ denotes the cardinality of $\mathcal{E}$ the set of edges in $\mathcal{G}$, is a convex function defined as $h(D) = \gamma\|D\|_1$. The proximal operator, where $\circ$ denotes the Hadamard product and $D = \nabla_{\mathcal{G}} x$ is

$$\text{prox}_{\tau_2 h}(D) = \text{sgn}(D) \circ \max(|D| - \tau_2\gamma, 0). \tag{3}$$

Using these tools, we can use the forward–backward-based primal dual approach presented in [39], for AGTV, to define Algorithm 1, where ϵ the stopping tolerance, I, J define the maximum number of iterations, and δ is a very small number to avoid a possible division by 0.

Complexity: We use the FLANN [38], whose computational complexity for n^2 patches of size l^2 each and fixed K is $\mathcal{O}(n^2 \log(n))$. Let J and I denote the maximum number of iterations for the algorithm to converge, then the computational cost of our algorithm is $\mathcal{O}(J|\mathcal{E}|I)$, where $|\mathcal{E}|$ denotes the number of nonzeros edges in the graph $\mathcal{G}$. For a $\mathcal{K}$-nearest neighbors graph, $|\mathcal{E}| \approx \mathcal{K}n^2$, so the computational complexity of our algorithm is linear in the size of the data sample n^2, i.e., $\mathcal{O}(J\mathcal{K}n^2 I)$. The graph $\mathcal{G}$ needs to be updated once in every outer iteration of the algorithm I, thus the overall complexity is $\mathcal{O}(IJ\mathcal{K}n^2 + In^2 \log(n))$. Dropping constants GTV scales with $\mathcal{O}(n^2)$ and AGTV scales with $\mathcal{O}(n^2(1 + \log(n))$.

Algorithm 1: Forward–backward primal dual for AGTV.

$x_0 = x_{\text{fbp}}$
1. INPUT: $U_0 = x_0$, $V_0 = \nabla_{\mathcal{G}} x_0$, $\epsilon > 0$
for $i = 0, \ldots I-1$ **do**
  for $j = 0, \ldots J-1$ **do**
    a. $P_j = \Phi(\text{prox}_{\tau_1 g}(\Phi^*(U_j) - \tau_1\Phi^*(\nabla_f(U_j) + \nabla^*_{\mathcal{G}} V_j))$
    b. $T_j = V_j + \tau_2 \nabla_{\mathcal{G}}(2P_j - U_j)$
    c. $Q_j = T_j - \tau_2 \,\text{prox}_{\frac{1}{\tau_2}h}\left(\frac{1}{\tau_2} T_j\right)$
    d. $(U_{j+1}, V_{j+1}) = (U_j, V_j) + \tau_3((P_j, Q_j) - (U_j, V_j))$
    if $\frac{\|U_{j+1}-U_j\|_F^2}{\|U_j\|_F^2+\delta} < \epsilon$ and $\frac{\|V_{j+1}-V_j\|_F^2}{\|V_j\|_F^2+\delta} < \epsilon$ **then**
      BREAK
    end if
  end for
  2. $x_i = U_{j+1}$
  3. Construct patch graph $\mathcal{G}$ from x_i
  if $\frac{\|x_i - x_{i-1}\|_F^2}{\|x_i\|_F^2+\delta} < \epsilon$ **then**
    BREAK
  end if
end for
OUTPUT: x_i

IV. Experimental Results

To test the performance of our AGTV method, we perform reconstructions for many different types of phantoms from different number of projections with varying levels of Poisson noise, using GSPBOX [41], UNLocBox [42], and AIRTools [16]. Reconstructions were judged on an ℓ_2 reconstruction error metric. We compare the performance of AGTV with many state-of-the-art iterative and convex optimization-based algorithms, which include FBP, ART (Kaczmarz), SIRT (Cimmino), CS, CSTV, and GTV (FLANN approximation of NLTV). All hyperparameters were tuned for best performance.

Each of these methods has its own model parameters, which need to be set or tuned in an appropriate manner. ART (Kaczmarz) and SIRT (Cimmino) were performed using FBP as *a priori*. The stopping criteria for ART and SIRT was set to 100 iterations and the relaxation parameter (η) was tuned to achieve the best result. For the graph-based reconstruction (GTV and AGTV), a graph prior $\mathcal{G}$ was generated by dividing the result from FBP into patches as explained previously. For example, for a Shepp–Logan phantom of size 64×64, the graph was constructed by dividing it into $64 \times 64 = 4096$ overlapping patches of size 3×3, $\mathcal{K} = 15$, and setting σ for the weight matrix to the average distance of the 15-nearest neighbors. For Algorithm 1, we set $I = J = 30$ and the convergence parameters τ_1, τ_2, τ_3 were set automatically by UNLocBox. It is worth mentioning here that GTV is a faster method of implementing NLTV by using $\mathcal{K}$-nearest neighbors graph approximation. Thus, the

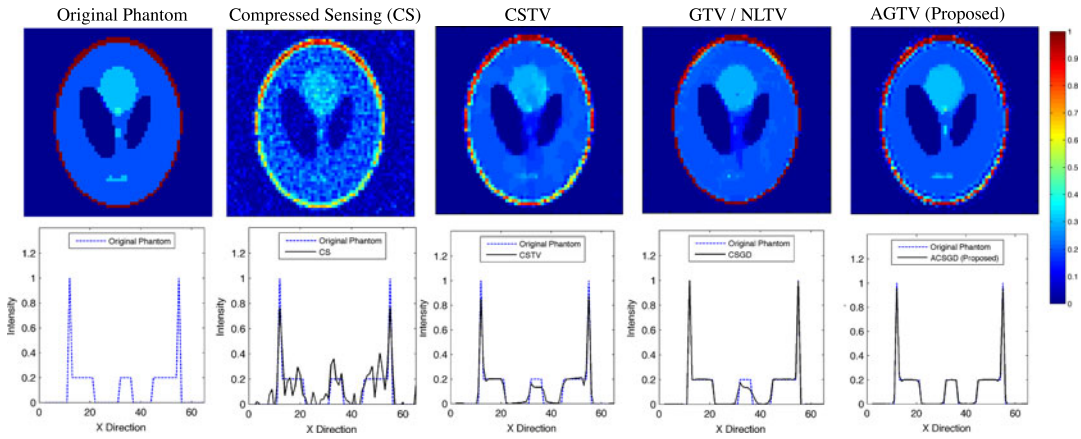

Fig. 2. Comparative analysis of reconstructing Shepp–Logan using various reconstruction methods. The sinogram of a 64 × 64 Shepp–Logan phantom corrupted with 10% Poission noise was reconstructed using FBP (linearly interpolated, cropped Ram–Lak filter), CSTV ($\lambda = 0.5$, $\gamma = 0.1$, prior: FBP, stopping criteria = 100 iterations), GTV/NLTV ($\lambda = 0.5$, $\gamma = 0.2$, prior: patch graph from FBP, stopping criteria = 100 iterations), and AGTV ($\lambda = 0.5$, $\gamma = 1$, prior: patch graph from FBP updated every iteration, I and J in Algorithm 1 set to 30). AGTV gives a better intensity profile as compared to all other methods while preserving the edges.

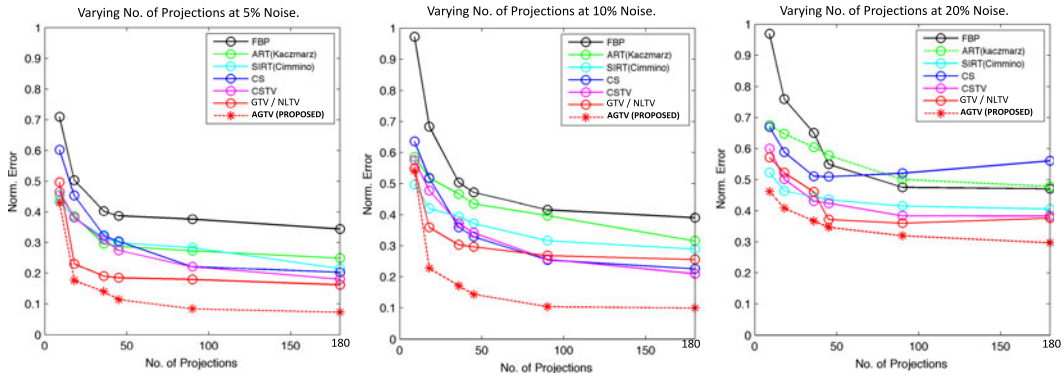

Fig. 3. Comparative analysis of reconstructing a Shepp–Logan phantom using various reconstruction methods at 5% and 10% Poisson noise. FBP (linearly interpolated, cropped Ram–Lak filter); ART (Kaczmarz/Randomized Kaczmarz, relaxation parameter (η) = 0.25, prior: FBP, stopping criteria = 100 iterations); SIRT (Cimmino/SART, (η) = 0.25, prior: FBP, stopping criteria = 100 iterations); CS (500 Iterations, prior: FBP); CSTV ($\lambda = 0.5$, $\gamma = 0.1$, prior: FBP, stopping criteria = 100 iterations); GTV ($\lambda = 0.5$, $\gamma = 0.2$, prior: patch graph from FBP, stopping criteria = 100 iterations); AGTV ($\lambda = 0.5$, $\gamma = 1$, prior: patch graph from FBP updated every iteration, I and J in Algorithm 1 set to 30).

GTV- and NLTV-based regularization are approximately equivalent in performance. Therefore, we did not include comparisons with NLTV.

To explain the performance of our model in detail, we reconstructed a 64 × 64 Shepp–Logan [43] phantom from 36 erroneous projections. A sinogram S was built by projecting the phantom using Radon transform and 36 equally spaced projections were collected from 0° to 180°. The sinogram was then corrupted with 10% Poission noise. Fig. 2, 3 provide a comparison of the reconstruction of the Shepp-Logan phantom using various reconstruction algorithms. It can be seen that AGTV performs better than GTV and CSTV. A similar experimental setup was repeated by reconstructing a 128 × 128 Torso phantom from 36 erroneous projections corrupted with 5% Gaussian normalized noise and similar results were achieved (Fig. 3 in the supplement). A graphical comparison for the reconstruction of Shepp–Logan using various reconstruction methods at varying number of projections and noise levels has been given in Fig. 3. AGTV shows promising results even with limited data reconstructions and outperforms many other state-of-the-art reconstruction and denoising methods. A more detailed analysis of these results has been presented in the supplement.

V. Conclusions

Similar to NLTV, our proposed method goes beyond spatial similarity between different regions of an image being reconstructed by establishing a connection between similar regions in the image regardless of spatial distance. However, it is much more scalable and computationally efficient because it uses the approximate nearest-neighbor search algorithm for graph construction, making it more likely to be adapted in a clinical setting. Beyond NLTV, our proposed approach is adaptive. The nonlocal graph prior is updated every iteration making the connection between similar regions stronger, thus, improving the overall reconstruction quality. Since TV is a special case of GTV, the proposed method can be seen as a generalization of CS and TV methods. Shortcomings of the proposed method include decreased graph quality due to approximations and tedious hyperparameter tuning.

REFERENCES

[1] J. Frank, *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*. New York, NY, USA: Springer Science & Business Media, 2008.

[2] A. Leis, M. Beck, M. Gruska, C. Best, R. Hegerl, and J. Leis, "Cryo-electron tomography of biological specimens," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 95–103, May 2006.

[3] A. Berrington de González, "Projected cancer risks from computed tomographic scans performed in the United States in 2007," *Arch. Intern. Med.*, vol. 169, no. 22, p. 2071, Dec. 2009.

[4] D. J. Brenner and E. J. Hall, "Computed tomography—An increasing source of radiation exposure," *N. Engl. J. Med.*, vol. 357, pp. 2277–2784, 2007.

[5] M. S. Pearce *et al.*, "Radiation exposure from ct scans in childhood and subsequent risk of leukaemia and brain tumours: A retrospective cohort study," *Lancet*, vol. 380, no. 9840, pp. 499–505, 2012.

[6] J. A. Fessler, "Statistical image reconstruction methods for transmission tomography," in *Handbook of Medical Imaging*, vol. 2, J. M. Fitzpatrick and M. Sonka, Eds. Bellingham, WA, USA: SPIE, 2000, pp. 1–70.

[7] Y. Censor, "Finite series-expansion reconstruction methods," *Proc. IEEE*, vol. 71, no. 3, pp. 409–419, Mar. 1983.

[8] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys. Med. Biol.*, vol. 51, no. 15, 2006, Art. no. R541.

[9] U. Skoglund, L.-G. Öfverstedt, R. M. Burnett, and G. Bricogne, "Maximum-entropy three-dimensional reconstruction with deconvolution of the contrast transfer function: A test application with adenovirus," *J. Struct. Biol.*, vol. 117, no. 3, pp. 173–188, Nov. 1996.

[10] H. Rullgård, O. Öktem, and U. Skoglund, "A componentwise iterated relative entropy regularization method with updated prior and regularization parameter," *Inverse Probl.*, vol. 23, no. 5, pp. 2121–2139, Oct. 2007.

[11] F. Natterer, *The Mathematics of Computerized Tomography*, vol. 32. Philadelphia, PA, USA: SIAM, 1986.

[12] E. T. Quinto, U. Skoglund, and O. Öktem, "Electron lambda-tomography," *Proc. Nat. Acad. Sci.*, vol. 106, no. 51, pp. 21 842–21 847, 2009.

[13] J. Hsieh, Computed Tomography: Principles, Design, Artifacts, and Recent Advances. Bellingham, WA, USA: SPIE, 2009.

[14] R. Gordon, R. Bender, and G. T. Herman, "Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theor. Biol.*, vol. 29, no. 3, pp. 471–481, Dec. 1970.

[15] G. Cimmino and C. N. delle Ricerche, *Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari*. Rome, Italy: Istituto per le applicazioni del calcolo, 1938.

[16] P. C. Hansen and M. Saxild-Hansen, "AIR tools—A MATLAB package of algebraic iterative reconstruction methods," *J. Comput. Appl. Math.*, vol. 236, no. 8, pp. 2167–2178, 2012.

[17] L. Landweber, "An iteration formula for Fredholm integral equations of the first kind," *Am. J. Math.*, vol. 73, no. 3, pp. 615–624, 1951.

[18] A. Brandt, "Algebraic multigrid theory: The symmetric case," *Appl. Math. Comput.*, vol. 19, no. 1, pp. 23–56, 1986.

[19] T. Strohmer and R. Vershynin, "A randomized kaczmarz algorithm with exponential convergence," *J. Fourier Anal. Appl.*, vol. 15, no. 2, pp. 262–278, 2009.

[20] C. G. Graff and E. Y. Sidky, "Compressive sensing in medical imaging," *Appl. Opt.*, vol. 54, no. 8, pp. C23–C44, 2015.

[21] 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," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, 2008.

[22] J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-ct," *Med. Phys.*, vol. 34, no. 11, pp. 4476–4483, 2007.

[23] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, "Improved total variation-based ct image reconstruction applied to clinical data," *Phys. Med. Biol.*, vol. 56, no. 6, p. 1545, 2011.

[24] J. Tang, B. E. Nett, and G.-H. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms," *Phys. Med. Biol.*, vol. 54, no. 19, pp. 5781–5804, Oct. 2009.

[25] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang, "Low-dose ct reconstruction via edge-preserving total variation regularization," *Phys. Med. Biol.*, vol. 56, no. 18, p. 5949, 2011.

[26] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, "Image recovery via nonlocal operators," *J. Sci. Comput.*, vol. 42, no. 2, pp. 185–197, 2010.

[27] G. Peyré, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2008, pp. 57–68.

[28] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model. Simul.*, vol. 7, no. 3, pp. 1005–1028, 2008.

[29] J. Huang and F. Yang, "Compressed magnetic resonance imaging based on wavelet sparsity and nonlocal total variation," in *Proc. 9th IEEE Int. Symp. Biomed. Imag.*, 2012, pp. 968–971.

[30] J. Liu, H. Ding, S. Molloi, X. Zhang, and H. Gao, "TICMR: Total image constrained material reconstruction via nonlocal total variation regularization for spectral CT," *IEEE Trans. Med. Imaging*, vol. 35, no. 12, pp. 2578–2586, Dec. 2016.

[31] X. Jia, Y. Lou, B. Dong, Z. Tian, and S. Jiang, "4D computed tomography reconstruction from few-projection data via temporal non-local regularization," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, Springer, 2010, pp. 143–150.

[32] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains," arXiv:1211.0053, 2012.

[33] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[34] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, Jul. 2017.

[35] O. Teke and P. Vaidyanathan, "Extending classical multirate signal processing theory to graphspart I: Fundamentals," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 409–422, Jan. 2017.

[36] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, Sep. 2014.

[37] F. Mahmood, N. Shahid, P. Vandergheynst, and U. Skoglund, "Graph-based sinogram denoising for tomographic reconstructions," in *Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2016, pp. 3961–3664.

[38] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2227–2240, Nov. 2014.

[39] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal? dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, Nov. 2015.

[40] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point Algorithms for Inverse Problems Science and Engineering*. Berlin, Germany: Springer, 2011, pp. 185–212.

[41] N. Perraudin, J. Paratte, D. Shuman, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," arXiv:1408.5781, 2014.

[42] N. Perraudin, D. Shuman, G. Puy, and P. Vandergheynst, "UNLocBox a matlab convex optimization toolbox using proximal splitting methods," arXiv:1402.0779, 2014.

[43] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. 21, no. 3, pp. 21–43, Jun. 1974.