Published in final edited form as:

Phys Med Biol. 2017 April 21; 62(8): 3284–3298. doi:10.1088/1361-6560/aa6392.

Assessment of Vectorial Total Variation Penalties on Realistic Dual-Energy CT Data

David S. Rigie, Adrian A. Sanchez, and Patrick J. La Rivière

University of Chicago, Department of Radiology

Abstract

Vectorial extensions of total variation have recently been developed for regularizing the reconstruction and denoising of multi-channel images, such as those arising in spectral computed tomography. Early studies have focused mainly on simulated, piecewise-constant images whose structure may favor total-variation penalties. In the current manuscript, we apply vectorial total variation to real dual-energy CT data of a whole turkey in order to determine if the same benefits can be observed in more complex images with anatomically realistic textures. We consider the total nuclear variation (TV_N) as well as another vectorial total variation based on the Frobenius norm (TV_F) and standard channel-by-channel total variation (TV_S). We performed a series of 3D TV denoising experiments comparing the three TV variants across a wide range of smoothness parameter settings, optimizing each regularizer according to a very-high-dose “ground truth” image. Consistent with the simulation studies, we find that both vectorial TV variants achieve a lower error than the channel-by-channel TV and are better able to suppress noise while preserving actual image features. In this real data study, the advantages are subtler than in the previous simulation study, although the TV_N penalty is found to have clear advantages over either TV_S or TV_F when comparing material images formed from linear combinations of the denoised energy images.

1. Background

Due to the recent progress in dual- and multi-energy x-ray CT, there has been growing interest in developing effective regularizers for multi-channel image reconstruction and denoising. In multi-energy scanning modes, the total x-ray dose is subdivided among two or more energy channels, thus reducing the signal-to-noise ratio (SNR) of each one. Furthermore, a pre- or post-reconstruction material decomposition step is usually performed, which greatly amplifies noise [1]. Some of these deleterious effects may be partially mitigated by including strong image priors into the reconstruction or post-processing steps. Several different approaches appear in the literature, including methods based on robust principal component analysis [2], patch-based penalties [3], and vectorial extensions of the total variation (TV) [4, 5, 6]. These varied techniques share a common goal of leveraging redundant structure across image channels (e.g. shared edges) in order to improve noise suppression while mitigating feature loss.

Previously, we applied a vectorial extension of the total variation penalty, which has been referred to as the “total nuclear variation” (TV_N), to spectral CT reconstruction [6]. For clarity, we will recap several salient points from that manuscript. TV_N extends the total

variation to multi-channel images by penalizing the nuclear norm of the Jacobian. The resulting functional encourages multi-channel images to have shared edges and gradient vectors that point in the same direction, while being insensitive to contrast or scale differences. Furthermore, TV_N is convex and reduces to the usual total variation when there is only a single image channel. In [6], we described a flexible image reconstruction framework based on a data-constrained TV minimization approach that allowed us to directly assess the performance of TV_N compared to a channel-by-channel application of the conventional TV. By jointly processing several image channels with edge-coupling regularization, improved reconstructions were demonstrated from the same noisy data in a series of simulation studies with the numerical XCAT phantom. Furthermore, the TV_N was applied to both energy and material images with similar gains relative to the channel-by-channel TV. Overall, TV_N allowed for a greater reduction in image noise with less edge blurring due to its ability to pool edge information across all image channels.

In this paper, we will build on those findings by investigating whether the same benefits can be demonstrated on more realistic CT data. In doing so, we aim to address several limitations of the previous study. Firstly, we will evaluate the TV_N on images that have a realistic texture. Although the XCAT phantom utilized in the previous study is based closely on real CT images of a human cadaver, its contrast is piecewise. This could be problematic for evaluating any sort of TV regularization, since TV is known to favor this kind of structure [7]. Furthermore, one of the potential benefits of the TV_N is that it promotes aligned gradient vectors across energy channels, but due to its piecewise nature, the XCAT phantom also trivially satisfies this condition. Other authors have applied a very similar regularizer to the joint reconstruction of PET-MR patient data with favorable results [8], but to our knowledge, no studies have reported on TV_N for real multi-energy CT images.

To address the question of whether the previous findings are translatable to realistic multi-energy CT data, we will investigate the proposed regularizer on dual-energy images of a turkey acquired on a diagnostic, clinical scanner. While the previous study focused on image reconstruction, the TV_N can also be applied to other image restoration problems [4, 5]. Since we did not have access to the projection data or geometry parameters of our scanner, we focus solely on denoising in this work. Because denoising is much less computationally expensive than reconstruction, this enabled us to conduct a thorough exploration of the parameter space, which was not feasible in the previous study. In addition to assessing the TV_N on realistic data, we also address several other issues of practical interest. Firstly, in order to isolate the importance of edge-coupling and gradient alignment, we also study another multi-channel variant of the TV [9] that encourages joint sparsity of the gradient magnitude images, without explicit encouragement of aligned gradient vectors. Additionally, we demonstrate that the TV_N can be efficiently extended to 3D image volumes, which is of practical relevance for reconstructing or denoising multi-detector and cone beam CT data.

2. Theory

2.1. Notation and definitions

In this section, we outline the notation that we will use for the remainder of the document. We adopt a somewhat unconventional set of indexing rules to maintain as much clarity as

possible in describing algebraic operations on vectors with multiple dimensions of spatial and spectral information.

Single-Channel images—Consider a discretized 3D image $u \in I$ where $I = \mathbb{R}^{M \times N \times P}$ is a finite dimensional vector space equipped with an inner product

$$\langle u, v \rangle_I = \sum_{i,j,k} u(i, j, k) v(i, j, k), \quad u, v \in I. \quad (1)$$

We use the convention $u(i, j, k)$ to refer to a particular voxel, where i, j , and k specify the row, column, and slice indices and $u(i, j, k) \in \mathbb{R}$ is a discretized version of some continuous image function $u(x, y, z)$. The integers M, N , and P specify the total number of rows, columns, and slices, respectively. The quantity ∇u is a vector in the space $G = I \times I$, where the operator $\nabla u : I \rightarrow G$ represents a discrete approximation to the gradient. At each voxel location we define the quantity $(\nabla u)(i, j, k) \in \mathbb{R}^3$ as

$$(\nabla u)(i, j, k) = \begin{pmatrix} (\nabla u)^x(i, j, k) \\ (\nabla u)^y(i, j, k) \\ (\nabla u)^z(i, j, k) \end{pmatrix}, \quad (2)$$

where

$$(\nabla u)^x(i, j, k) = \begin{cases} u(i, j, k+1) - u(i, j, k) & k < P \\ 0 & k = P \end{cases} \quad (3)$$

$$(\nabla u)^y(i, j, k) = \begin{cases} u(i, j+1, k) - u(i, j, k) & j < N \\ 0 & j = N \end{cases} \quad (4)$$

$$(\nabla u)^z(i, j, k) = \begin{cases} u(i+1, j, k) - u(i, j, k) & i < M \\ 0 & i = M \end{cases}. \quad (5)$$

We also define an inner product in G ,

$$\langle \nu, z \rangle_G = \sum_{\alpha \in \{x, y, z\}} \sum_{i, j, k} \nu^\alpha(i, j, k) z^\alpha(i, j, k), \quad \nu, z \in G. \quad (6)$$

Note that we use bold font to indicate that each spatial location (i, j, k) maps to a vector of gradient values. Further, we will need a discrete divergence operator $\text{div } \mathbf{z} : G \rightarrow I$, which is chosen to be the negative transpose of the gradient operator, defined by

$$\langle \nabla u, \mathbf{z} \rangle_G = -\langle u, \text{div } \mathbf{z} \rangle_I. \quad (7)$$

Lastly, we define the mixed ℓ_1/ℓ_2 -norm in G as

$$\|\mathbf{z}\|_{1,2} = \sum_{i,j,k} \|\mathbf{z}(i, j, k)\|_2 \quad \mathbf{z} \in G, \quad (8)$$

indicating that we take an ℓ_1 -norm over the spatial indices (i, j, k) and an ℓ_2 -norm of each 3-vector $\mathbf{z}(i, j, k)$. This mixed-norm notation is often used in the literature on sparse regression and occasionally to compactly define the isotropic total variation [10],

$$\text{TV}(u) = \|\nabla u\|_{1,2}. \quad (9)$$

Multi-channel images—Now, we consider a discrete image $\mathbf{u} \in \mathcal{J}$ with L spectral channels,

$$\mathbf{u}(i, j, k) = \begin{pmatrix} u_1(i, j, k) \\ u_2(i, j, k) \\ \vdots \\ u_L(i, j, k) \end{pmatrix}, \quad (10)$$

where $\mathcal{J} = \mathbb{R}^{L \cdot M \cdot N \cdot P}$. The quantity $\mathbf{u}$ is a discretized version of some continuous vector field $\mathbf{u}(x, y, z)$, such as a multi-energy image volume. We also define an inner product in $\mathcal{J}$,

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{J}} = \sum_{\ell} \sum_{i,j,k} u_{\ell}(i, j, k) v_{\ell}(i, j, k) \quad \mathbf{u}, \mathbf{v} \in \mathcal{J}, \quad (11)$$

where the subscript ℓ denotes a particular image channel. Since $\mathbf{u}$ is the discrete analog of a vector field, we can also define the discrete Jacobian, $\mathbf{J}\mathbf{u} : \mathcal{J} \rightarrow \mathcal{G}$, which generalizes the gradient operator to vector fields. In particular we have

$$(J\mathbf{u})(i, j, k) = \begin{pmatrix} (\nabla \mathbf{u}_1)^x(i, j, k) & (\nabla \mathbf{u}_1)^y(i, j, k) & (\nabla \mathbf{u}_1)^z(i, j, k) \\ (\nabla \mathbf{u}_2)^x(i, j, k) & (\nabla \mathbf{u}_2)^y(i, j, k) & (\nabla \mathbf{u}_2)^z(i, j, k) \\ \vdots & \vdots & \vdots \\ (\nabla \mathbf{u}_L)^x(i, j, k) & (\nabla \mathbf{u}_L)^y(i, j, k) & (\nabla \mathbf{u}_L)^z(i, j, k) \end{pmatrix}. \quad (12)$$

At every voxel (i, j, k) , the $L \times 3$ sub-matrix $(J\mathbf{u})(i, j, k)$ fully characterizes the first-order derivatives of $\mathbf{u}$, with each row consisting of the gradient vector of one of the L image channels. The quantity $J\mathbf{u}$ is in vector space $\mathcal{G} = \mathcal{I} \times \mathcal{J} = \mathbb{R}^{(L \times 3) \times (M \times N \times P)}$ with an inner product

$$\langle V, Z \rangle_{\mathcal{G}} = \sum_{\alpha \in \{x, y, z\}} \sum_{\ell} \sum_{i, j, k} V_{\ell}^{\alpha}(i, j, k) Z_{\ell}^{\alpha}(i, j, k) \quad V, Z \in \mathcal{G}. \quad (13)$$

An element $V \in \mathcal{G}$ is a discretized version of a tensor field $V(x, y, z)$, so we use uppercase font to indicate that every spatial location (i, j, k) maps to a matrix. Again, we need an analog of the divergence operator $\text{Div } Z: \mathcal{G} \rightarrow \mathcal{J}$ that is the negative transpose of J . It is constructed to satisfy

$$\langle J\mathbf{u}, Z \rangle_{\mathcal{G}} = -\langle \mathbf{u}, \text{Div } Z \rangle_{\mathcal{J}}. \quad (14)$$

We define the mixed ℓ_1 /nuclear-norm as

$$\|Z\|_{1, \star} = \sum_{i, j, k} \|Z(i, j, k)\|_{\star} \quad Z \in \mathcal{G}, \quad (15)$$

where $\|Z(i, j, k)\|_{\star}$ is referred to as the “nuclear norm” of matrix $Z(i, j, k)$ and is equal to the sum of its singular values. Additionally, we define the mixed ℓ_1 /Frobenius norm as

$$\|Z\|_{1, F} = \sum_{i, j, k} \|Z(i, j, k)\|_F \quad Z \in \mathcal{G}, \quad (16)$$

where the Frobenius norm of a generic $M \times N$ matrix is given by

$$\|X\|_F = \sqrt{\sum_{i=1}^M \sum_{j=1}^N X_{ij}^2}. \quad (17)$$

Denoising Model—In this section, we will describe a generic denoising model as well as the three variants of the TV that we have investigated. Our denoising model is specified by the following optimization problem:

$$\operatorname{argmin}_u \quad \frac{1}{2} \|u - g\|_2^2 + \lambda \operatorname{TV}(u), \quad (18)$$

where g is the noisy image, λ controls the trade-off between data fidelity and smoothness, and TV refers generally to some variant of the total variation. The simplest way to generalize the TV to multi-channel images is to apply the standard isotropic TV separately to each energy channel; we refer to this as the channel-by-channel TV, denoted by TV_S . Additionally, we can define a “vectorial” TV that simultaneously penalizes edges across all image channels, which is defined by the voxel-wise application of some matrix norm to the discrete Jacobian of the multi-channel image function. Two interesting choices are the Frobenius norm, which encourages joint sparsity in the image gradients, and the Nuclear norm, which encourages joint sparsity *and* directional alignment of gradient vectors [5, 6]. We denote these two vectorial TV variants by TV_F and TV_N , respectively. All three of these regularizers are summarized in Table 1.

Primal-Dual Saddlepoint Problem—As in [6], the denoising model can be expressed as a saddlepoint problem, which is given by

$$\min_u \max_Z \quad \langle Ju, Z \rangle_{\mathcal{G}} - \delta_{\mathcal{K}}(Z) + \frac{1}{2\lambda} \|u - g\|_2^2, \quad (19)$$

where $\delta_{\mathcal{K}}$ is a generic set indicator function defined by

$$\delta_{\mathcal{K}}(\mathbf{X}) \equiv \begin{cases} 0 & \mathbf{X} \in \mathcal{K} \\ \infty & \mathbf{X} \notin \mathcal{K} \end{cases}. \quad (20)$$

The set $\mathcal{K}$ is equal to $\mathcal{S}$, $\mathcal{F}$, or $\mathcal{N}$ for TV_S , TV_F , and TV_N , respectively. The sets $\mathcal{S}$, $\mathcal{F}$, and $\mathcal{N}$ are defined below:

$$\mathcal{S} = \{Z \in \mathcal{G} : \|Z_{\ell}(i, j, k)\|_2 \leq 1 \forall i, j, k, \ell\} \quad (21)$$

$$\mathcal{F} = \{Z \in \mathcal{G} : \|Z(i, j, k)\|_F \leq 1 \forall i, j, k\} \quad (22)$$

$$\mathcal{N} = \{Z \in \mathcal{G} : \sigma_{\max}(Z(i, j, k)) \leq 1 \forall i, j, k\}. \quad (23)$$

Algorithm Updates—As in [6], we solve this optimization problem using Chambolle and Pock's first-order, primal-dual algorithm (CPPD) [11]. However, unlike the reconstruction problem described in [6], this saddle point problem is strongly convex in the primal variable $\mathbf{u}$, with convexity constant λ^{-1} , so we can use their accelerated algorithm. This version of CPPD achieves a convergence rate of $O(1/N^2)$ rather than $O(1/N)$. The updates are given by Algorithm 1.

The primal and dual stepsizes τ and σ must be initialized so that $\tau\sigma\|J\|^2 = 1$, where $\|J\|$ is the spectral radius (also the maximum singular value) of the Jacobian operator J . For 3D image arrays, It can be shown that $\|J\| \leq \sqrt{12}$ [12]. We find that convergence is always fastest when $\sigma\tau = 1/12$; however there still exists freedom to choose the ratio of primal to dual stepsizes. The observed rate of convergence can be dramatically impacted by this choice, but we are unaware of any good method to automatically select these parameters. We have used $\tau = 0.05$ for all of the results in this paper. We iterate until the primal-dual gap falls below 10^{-6} , which we have found, empirically, to be a reasonable stopping point where no appreciable voxel value changes can be observed in the windows of interest.

Algorithm 1

TV Denoising Update Equations

```

1: Initialize:  $\sigma_k \tau_k = 1/12$ 
2: repeat
3:    $Z^{(k+1)} = \Pi_{\mathcal{Z}}(Z^{(k)} + \sigma_k J^{-1} \nabla \cdot (\lambda \mathbf{u}^{(k)} + \tau_k \mathbf{g}))$ 
4:    $\mathbf{u}^{(k+1)} = (\lambda \mathbf{u}^{(k)} + \tau_k \mathbf{g} + \lambda \tau_k \text{Div } Z^{(k+1)}) / (\lambda + \tau_k)$ 
5:    $\theta_k = \sqrt{\lambda / (\lambda + 2\tau_k)}$ ,  $\tau_k = \theta_k \tau_k$ ,  $\sigma_k = \sigma_k / \theta_k$ 
6:    $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k+1)} + \theta_k (\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)})$ 
7: until  $\triangleright$  stop when convergence criteria met

```

Implementation of Projection Operators—The algorithm updates involve projecting the dual variable Z onto one of the convex sets, $\mathcal{S}$, $\mathcal{F}$, or $\mathcal{N}$, corresponding to TV_S , TV_F , or TV_N . Pseudocode for each of these projection operators is given by Algorithm 2, Algorithm 3, and Algorithm 4, respectively. For TV_S and TV_F , the projection operators are very straightforward and only involve simple conditionals and vector normalization. The projection operator $\Pi_{\mathcal{N}}$ for TV_N is slightly more complicated and requires the computation of eigenvalues and eigenvectors of a real, symmetric 3×3 matrix for every image voxel. Luckily, analytic methods exist for diagonalizing 3×3 Hermitian matrices, so the computational burden is not too significant; we rely on the implementation described by Kopp [13]. We also note that for the special case of only two energy channels, there is an alternative implementation of $\Pi_{\mathcal{N}}$ that only involves a 2×2 eigendecomposition, but we

present this version because it generalizes to an arbitrary number of image channels, which is important for photon-counting spectral CT systems.

Algorithm 2

Implementation of Projection Operator $\Pi_{\mathcal{S}}(Z)$

```

1: for all  $i, j, k, \ell$  do
2:    $x \leftarrow Z_{\ell}(i, j, k)$ 
3:   if  $\|x\|_2 = 1$  then
4:      $x \leftarrow x/\|x\|_2$ 
5:   end if
6:    $Z_{\ell}(i, j, k) \leftarrow x$ 
7: end for

```

Algorithm 3

Implementation of Projection Operator $\Pi_{\mathcal{C}}(Z)$

```

1: for all  $i, j, k$  do
2:    $X \leftarrow Z(i, j, k)$ 
3:   if  $\|X\|_F = 1$  then
4:      $X \leftarrow X/\|X\|_F$ 
5:   end if
6:    $Z(i, j, k) \leftarrow X$ 
7: end for

```

3. Denoising Experiments

Data Acquisition—A series of scans were acquired of a frozen supermarket turkey at 80 and 140 kVp on a 256-slice multidetector CT (Brilliance iCT, Philips Healthcare) in axial mode. At each kVp, we acquired both a “low-dose” scan (100mAs) and ten high dose scans (410 mAs for 80 kVp and 425 mAs for 120/140 kVp). The high-dose scans were repeated so that the resulting images could be averaged to create a very low-noise “ground truth” image. All 33 scans were acquired with the exact same table position and no other parameters were altered between acquisitions besides the kVp and mAs.

Algorithm 4

Implementation of Projection Operator $\Pi_{\mathcal{N}}(Z)$

```

1: for all  $i, j, k$  do
2:    $X \leftarrow Z(i, j, k)$ 
3:    $(\lambda, V) \leftarrow \text{Eig}(X^T X) \triangleright$  compute eigenvalues and eigenvectors [13]
4:    $\Sigma \leftarrow \text{diag}(\sqrt{\lambda})$ 
5:   if  $\text{Tr}(\Sigma) > 1$  then

```

```

6:       $\Sigma_p \leftarrow \min(1, \Sigma) \triangleright$  element-wise minimum
7:       $X \leftarrow XV\Sigma + \Sigma_p V^T$ 
8:      end if
9:       $Z(i, j, k) \leftarrow X$ 
10: end for

```

All of the scans were reconstructed with FBP (high resolution mode, Filter E) via the manufacturer's software and using the exact same parameters: $512 \times 512 \times 128$ image volumes with 0.562mm voxels and a 0.625mm slice thickness. Linear interpolation was used along the z-axis (axial direction) to synthesize a volume with isotropic resolution (0.562mm) to avoid complications with the discrete gradient and divergence operators. Since the scan was acquired in axial mode, the slices toward the axial extremes of the acquired data volume exhibited severe cone beam artifacts. A smaller volume consisting of only the central 32 slices was extracted in order to minimize these effects.

Comparison of TV_S , TV_F , and TV_N —In order to compare these TV variants, we synthesized a dual kVp image pair by extracting the same 3D volume from the 80 and 140 kVp low-dose (100mAs) datasets. Then we denoised this dual-energy data over a wide range of λ values with each of TV_S , TV_F , and TV_N ; we first identified $\lambda_{\min}$ and $\lambda_{\max}$, corresponding to images that appeared under- and over-smoothed and sampled 100 equally spaced values within this interval. Each result was then compared to the corresponding 10X averaged high-dose image, and the normalized Euclidean distance was computed as an error metric. This is defined as

$$\varepsilon(\lambda) = \frac{\|u^*(\lambda) - u_{\text{ref}}\|_2}{\|u_{\text{ref}}\|_2}, \quad (24)$$

where $u^*(\lambda)$ is the denoised image with parameter λ , and u_{ref} is the corresponding 10X high-dose reference image. In the case of TV_F and TV_N , the dual-energy data are processed jointly due to the interchannel coupling of the regularizer, but we report the error metrics individually for each channel. This is shown in Figure 1.

In general, the error curves should not be compared at any one specific common value of λ because the functions TV_S , TV_F , and TV_N have inherently different scales. Instead the curves should be evaluated in terms of the possible family of solutions they each represent and, most importantly, their minimum-error solution. As can be seen in Figure 1, the channel-coupling of the VTV has little effect on the 140 kVp image; however, it does improve the accuracy of the noisier 80 kVp image, where it is clear that TV_N outperforms TV_F , which in turn outperforms TV_S . The minima of the various error curves is reported for both the 80 and 140 kVp channels for all regularizers in Table 2. In addition to relative error, which can be thought of as a normalized root-mean-square error (RMSE), we also report the universal quality index (UQI) proposed in [14]. For two images x and y , the UQI is defined as

$$UQI = \frac{4\sigma_{xy}\overline{xy}}{(\sigma_x^2 + \sigma_y^2)[(\overline{x})^2 + (\overline{y})^2]}, \quad (25)$$

where

$$\begin{aligned} \overline{x} &= \frac{1}{N} \sum_{i=1}^N x_i, & \overline{y} &= \frac{1}{N} \sum_{i=1}^N y_i \\ \sigma_x^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \overline{x})^2, & \sigma_y^2 &= \frac{1}{N-1} \sum_{i=1}^N (y_i - \overline{y})^2 \\ \sigma_{xy} &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \overline{x})(y_i - \overline{y}). \end{aligned}$$

In particular, the UQI is computed between the high dose reference images and the denoised images and reported in Table 2 in parenthesis. The dynamic range of UQI is $[-1, 1]$, where higher values are better.

Figure 2 shows a comparison between the denoised 80 kVp images between different VTV variants, each shown at its optimal lambda value (the minimizer of the error curve); Figure 3 shows the same comparison for an ROI. It can be seen that the channel-coupling does improve the preservation of details and suppression of noise. The differences between TV_F and TV_N are quite subtle with TV_N arguably having slightly cleaner edges, but the improvement over channel-by-channel TV (TV_S) is somewhat more pronounced. Figures 4 and 5 shows the corresponding results for the 140 kVp channel. In that case, differences between regularizers are practically indiscernible. This observation is consistent with our previous findings that the channel-coupling effect of these VTV penalties is primarily helpful to the noisier image channels because it allows edge information to be transferred from the less noisy channel.

4. Water-Subtracted Images

In principle, one can use dual-energy CT measurements to solve for quantitative material density maps; however, this sort of material decomposition requires either accurate knowledge of the input spectrum or a significant amount of calibration data. Alternatively, one can simply take linear combinations of the reconstructed 80 and 140 kVp images, where the coefficients are chosen to cancel out a given material [15]. For example, merely subtracting the 140 kVp image from the 80 kVp image should approximately cancel out any voxels that have an effective atomic number very similar to water, provided that the original images have already been converted to Hounsfield units (HU). Though the resultant image lacks quantitative accuracy, it is qualitatively very similar to a bone density image.

After denoising the low-dose 80/140 kVp images with the various regularizers (TV_S , TV_F , and TV_N), we formed water-subtracted images from each image pair. We find that the

differences between the various regularizers is far more pronounced in this water-subtracted image than in either the 80 or 140 kVp images, which can be seen in figures 7 and 8. Compared with TV_S , TV_N does a better job at preserving fine details in the bony structures while also achieving a less noisy texture. Visually, TV_N also appears to slightly outperform TV_F in this regard, indicating that the effect of encouraging aligned gradient vectors across image channels may be important, despite the fact that TV_F and TV_N barely show any visible differences at all in the raw 80/140 kVp images. These differences are quantified and confirmed in table 3, wherein the contrast-to-noise ratio (CNR) is computed between an inclusion in the vertebral structure and the nearby bone for each water-subtracted image. We define CNR as

$$CNR = \sqrt{2 \frac{(\overline{HU}_{bone} - \overline{HU}_{inclusion})^2}{\sigma_{bone}^2 + \sigma_{inclusion}^2}}, \quad (26)$$

where $\overline{HU}_{inclusion}$ is the mean HU in the ROI shown in figure 6 and $\overline{HU}_{bone}$ is an identically sized ROI placed over bone. $\sigma_{inclusion}^2$ and σ_{bone}^2 are the variances in these ROIs.

As in Table 2, we also report the relative error and UQI metrics for the water-subtracted images in Table 4.

5. Discussion

In this paper, we assessed the utility of the vectorial TV for realistic CT images. Consistent with [6], we found that the edge-coupling properties of TV_N can help to preserve details and also improve quantitative accuracy, compared with channel-by-channel TV. The advantages were most striking when examining material images formed as linear combinations of the denoised energy images. In this study, the benefits were more subtle than in the reconstruction study utilizing the numerical XCAT phantom. This is not particularly surprising since that phantom is piecewise constant, which makes it possible to operate in a regime of very strong regularization.

In addition, we looked at TV_F , which also encourages common edges but is ambivalent to gradient-vector alignment between image channels. Although TV_N led to more accurate images (relative to the reference images) and arguably cleaner edges, especially for material images, TV_F has some practical advantages, as it is simpler to implement and slightly less computationally expensive. However, we have demonstrated that TV_N can be implemented efficiently in 3D for an arbitrary number of energy channels using the analytic matrix diagonalization method described in [13]. For this denoising study, TV_N involves slightly longer computation times than TV_F or TV_S ($\approx 20\%$ longer run time in our implementation), but this difference is negligible for image reconstruction because the the projection and backprojection steps are the dominant factors determining the per-iteration runtime.

Interestingly, the benefits of the inter-channel coupling were most apparent when there was a very large disparity in the noise levels between the low- and high-energy image channels.

Intuitively, this makes sense because both TV_F and TV_N allow for edge information to be transferred from the low-noise image into noisy image. This could be useful for making multi-energy CT acquisitions more robust to scenarios where certain channels receive low count rates, such as dual kVp imaging of obese patients.

6. Conclusion

The edge-coupling properties of TV_F and TV_N lead to improved image quality and accuracy compared to the conventional total variation when denoising multi-channel images. On real data, with realistic anatomical texture and appearance, the improvement is not as pronounced as it was in previous numerical experiments with piecewise constant phantoms[6]. However, TV_N consistently outperforms TV_F and TV_S in terms of root-mean-square error, universal quality index, and subjective visual assessment.

Acknowledgments

This work was supported in part by Toshiba Medical Research Institute USA, NIH grant R01EB017293, the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under grant number T32 EB002103, and the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R25GM109439 (Project Title: University of Chicago Initiative for Maximizing Student Development (IMSD)). Partial funding for the computation in this work was provided by NIH Grant Nos. S10 RRO21039 and P30 CA14599. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or Toshiba Medical Research Institute USA.

References

1. Alvarez RE. Med Phys. 2013; 40:111909. [PubMed: 24320442]
2. Gao H, Yu H, Osher S, Wang G. Inverse Probl. 2011; 27
3. Kim K, Ye JC, Worstell W, Ouyang J, Rakvongthai Y, El Fakhri G, Li Q. IEEE Trans Med Imaging. 2015; 34:748–760. [PubMed: 25532170]
4. Lefkimmatis, S., Roussos, A., Unser, M., Maragos, P. 2013 Convex generalizations of total variation based on the structure tensor with applications to inverse problems. In: Kuijper, A. Bredies, K. Pock, T., Bischof, H., editors. Scale Space and Variational Methods in Computer Vision. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 48–60. (Lecture Notes in Computer Science vol 7893)
5. Holt K. IEEE Trans Image Process. 2014; 23:3975–3989.
6. Rigie DS, La Rivière PJ. Phys Med Biol. 2015; 60:1741–1762. [PubMed: 25658985]
7. Chan T, Marquina A, Mulet P. SIAM J Sci Comput. 2000; 22:503–516. (Preprint <http://dx.doi.org/10.1137/S1064827598344169>).
8. Knoll F, Koesters T, Otazo R, Block T, Feng L, Vunckx K, Faul D, Nuyts J, Boada F, Sodickson DK. EJNMMI Phys. 2014; 1:A26. [PubMed: 26501612]
9. Bresson X, Chan TF. Inverse Probl Imaging. 2008; 2:455–484.
10. Andersen MS, Hansen PC. Numer Algorithms. 2013; 67:121–144.
11. Chambolle A, Pock T. J Math Imaging Vis. 2010; 40:120–145.
12. Beck A, Teboulle M. IEEE Trans Image Process. 2009; 18:2419–2434. [PubMed: 19635705]
13. Kopp J. International Journal of Modern Physics C. 2008; 19:523–548. URL <http://www.worldscientific.com/doi/abs/10.1142/S0129183108012303>.
14. Wang Z, Bovik AC. IEEE Signal Process Lett. 2002; 9:81–84.
15. Lehmann LA, Alvarez RE, Macovski A, Brody WR, Pelc NJ, Riederer SJ, Hall AL. Med Phys. 1981; 8:659–667. [PubMed: 7290019]

Figure 1.

Relative error curves based on (24) are shown for the three different VTV variants studied at both 80 kVp (left) and 140 kVp (right). Optimal values of λ are identified by the minima of these curves.

Figure 2.

Reconstructed 80 kVp images corresponding to low-dose data (a), 10x averaged high-dose data (b) and denoised low-dose images with TV_S , TV_F , and TV_N regularization, respectively (c–e). The parameter λ is selected to optimize each regularizer according to figure 1, and the display window is $[-200 \text{ HU}, 200 \text{ HU}]$.

Figure 3. ROI from reconstructed 80 kVp images corresponding to low-dose data (a), 10x averaged high-dose data (b) and denoised low-dose images with TV_S , TV_F , and TV_N regularization, respectively (c–e). The parameter λ is selected to optimize each regularizer according to figure 1, and the display window is $[-200 \text{ HU}, 200 \text{ HU}]$.

Figure 4.

Reconstructed 140 kVp images corresponding to low-dose data (a), 10x averaged high-dose data (b) and denoised low-dose images with TV_S , TV_F , and TV_N regularization, respectively (c–e). The parameter λ is selected to optimize each regularizer according to figure 1, and the display window is $[-200 \text{ HU}, 200 \text{ HU}]$.

Figure 5. ROI from reconstructed 80 kVp images corresponding to low-dose data (a), 10× averaged high-dose data (b) and denoised low-dose images with TV_S , TV_F , and TV_N regularization, respectively (c–e). The parameter λ is selected to optimize each regularizer according to figure 1, and the display window is $[-200 \text{ HU}, 200 \text{ HU}]$.

Figure 6.
The ROI placed over the inclusion is depicted by the yellow rectangle.

Figure 7. Water-subtracted images corresponding to low-dose data (a), 10 $\times$ averaged high-dose data (b) and denoised low-dose images with TV_S , TV_F , and TV_N regularization, respectively (c–e). The parameter λ is selected to optimize each regularizer according to figure 1.

Figure 8. ROI from water-subtracted kVp images corresponding to low-dose data (a), 10× averaged high-dose data (b) and denoised low-dose images with TV_S , TV_F , and TV_N regularization, respectively (c–e). The parameter λ is selected to optimize each regularizer according to figure 1.

Table 1

Three multi-channel variants of the TV studied in this work. Note, TV_S induces no inter-channel coupling and is equivalent to separately denoising each image channel.

Defintion	Edge Coupling	Gradient Alignment
$TV_S(\mathbf{u}) = TV_S(\mathbf{u}) = \sum_{\ell=1}^L TV(u_\ell)$		
$TV_F(\mathbf{u}) = \ \mathcal{J}\mathbf{u}\ _{1,F}$	✓	
$TV_N(\mathbf{u}) = \ \mathcal{J}\mathbf{u}\ _{1,\star}$	✓	✓

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Relative error and UQI (in parenthesis) of each VTV variant at its respective optimal λ value. These correspond to the minima of the error curves in Figure 1, $\varepsilon(\lambda_{\text{opt}})$.

Regularizer	80 kVp	140 kVp
TV _S	5.501e-2 (0.9981)	3.101e-2 (0.9994)
TV _F	4.820e-2 (0.9985)	3.113e-2 (0.9994)
TV _N	3.876e-2 (0.9990)	3.027e-2 (0.9994)

Table 3

A comparison of the contrast-to-noise ratio (CNR) between an inclusion in the vertebral structure and the adjacent solid bone across TV_S , TV_F , and TV_N in the water-subtracted image.

Image	CNR
Low Dose (Original)	2.4
10× Dose (Reference)	6.0
TV_S	3.0
TV_F	3.2
TV_N	5.7

Table 4

Relative error and UQI of each VTV variant at its respective optimal λ value for the water-subtracted images.

Regularizer	Relative Error	Universal Quality Index (UQI)
TV _S	0.6260	0.8044
TV _F	0.3768	0.8885
TV _N	0.3261	0.9258