

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

Image reconstruction for view-limited x-ray CT in baggage scanning

Sagar Mandava, David Coccarelli, Joel A. Greenberg,
Michael E. Gehm, Amit Ashok, et al.

SPIE.

Image reconstruction for view-limited X-ray CT in baggage scanning

Sagar Mandava^a, David Coccarelli^b, Joel A. Greenberg^b, Michael E. Gehm^b, Amit Ashok^{a,c},
and Ali Bilgin^a

^aDepartment of Electrical and Computer Engineering, University of Arizona, Tucson, AZ, USA

^bDepartment of Electrical and Computer Engineering, Duke University, Durham, NC, USA

^cCollege of Optical Sciences, University of Arizona, Tucson, AZ, USA

ABSTRACT

X-ray CT based baggage scanners are widely used in security applications. Recently, there has been increased interest in view-limited systems which can improve the scanning throughput while maintaining the threat detection performance. However as very few view angles are acquired in these systems, the image reconstruction problem is challenging. Standard reconstruction algorithms such as the filtered backprojection create strong artifacts when working with view-limited data. In this work, we study the performance of a variety of reconstruction algorithms for both single and multi-energy view-limited systems.

Keywords: Image reconstruction, security scanning, view-limited imaging

1. INTRODUCTION

X-ray computed tomography (CT) is an imaging modality that is commonly used in medical imaging. It has also become an integral part of security applications, especially for baggage and cargo screening. A major difference between CT imaging for medical applications as opposed to security applications is that the former deals with a limited class of objects while the object diversity in security applications is virtually unconstrained. Both applications also have different operational constraints. One of the primary concerns in medical CT imaging is the dose received by the patient since X-rays are ionizing radiation. On the other hand, security imaging is primarily concerned with scanning throughput while maintaining the required threat detection rates.

In a simplified sense, scanning throughput is predominantly a function of the belt speed and the number of projections (views) being acquired. The belt speed is typically limited by the motion blur that can be tolerated for the application assuming enough projections are acquired to reconstruct the image. One strategy to improve throughput would be to simply increase the belt speed and use fewer projections to reconstruct the object being scanned. However, this approach would lead to significant artifacts when standard reconstruction techniques are used. Therefore, more sophisticated reconstruction algorithms are necessary to obtain images from view-limited data.

In this paper, we study the performance of several reconstruction algorithms on view-limited fan-beam CT data. We propose sparsity and rank constrained algorithms and compare them to a method that is a surrogate for standard CT reconstruction. A prototype fan-beam CT scanner with multi-energy resolution capability was used to scan a collection of custom built bags with a variety of objects that simulate typical baggage scanning applications. The acquired data was used to simulate single and dual energy systems corresponding to different number of views for each bag. Using this data, we investigate the performances of different reconstruction methods across view-reduction factors for single vs. dual energy systems.

Further author information: (Send correspondence to Sagar Mandava)

S. Mandava: E-mail: sagar88@email.arizona.edu

A. Ashok: E-mail: ashoka@email.arizona.edu

A. Bilgin: E-mail: bilgin@email.arizona.edu

2. METHODS

2.1 Data acquisition

Data was acquired on a prototype fan-beam CT scanner with the ability to resolve the signal along the spectral dimension into multiple energy channels.¹ The system used a detector with 1536 pixels of size 0.8mm and a field of view of 750mm. The spectral dataset consisted of 64 energy channels of which the first 48 were used to avoid the significant SNR drop seen in the last 16 channels. By merging the 48 channels into a single bin, a single energy system was simulated. A dual energy system was also simulated by merging the first 16 spectral channels to form the low energy spectral bin and the remaining 32 channels to form the high energy spectral bin. The raw data was acquired with an angular resolution of 0.02° over a 360° rotation. This data was retrospectively merged to generate a dataset with 720 angular views to boost SNR. The resulting dataset served as the reference dataset and was subsampled uniformly in the angular dimension to simulate view-limited acquisitions.

A collection of objects was created and packed using a predetermined ordering scheme to simulate bags representative of ones seen in typical baggage scanning applications. To create a bag, a subset of the available objects are arranged according to a specified configuration and this allows us to reuse the objects and allows the acquisition of a large number of bags. These bags contain objects that are of interest from a detection stand-point and several other objects that are meant to represent clutter. On this acquisition setup, data was acquired from 100 different bags which are used in this work. The images were reconstructed on a 375×375 pixel grid. A first pass pruning of non-ideal detector pixels was performed to minimize ring artifacts. Sample bags illustrating the diversity of objects and the variety of clutter in these experiments are shown in Figure 1.

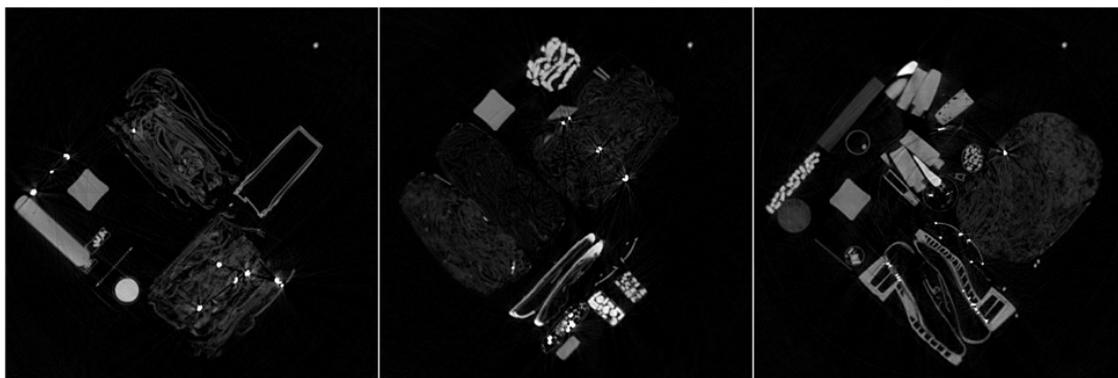


Figure 1. Sample bags with different levels of clutter seen on the high energy bin of a dual energy system. From left to right, the amount of clutter in the scene is increasing.

2.2 Image reconstruction

2.2.1 Data model

In a linear model for X-ray CT, the projection data ($y \in \mathbb{R}^M$) and the underlying object ($x \in \mathbb{R}^N$) are related by the forward projection operator (A) as follows:

$$y = Ax + n, \quad (1)$$

where $n \in \mathbb{R}^M$ is a noise term.

The operator A can be represented in matrix form ($A \in \mathbb{R}^{M \times N}$) and is often referred to as the system matrix. The system matrix models the geometry of the CT system. In this work, the system matrix was created using routines available in the Image Reconstruction Toolbox.² The source to detector and iso-center to detector distances, detector pixel sizes were carefully calibrated and provided as inputs to generate the system matrix. The detector pixel array is a flat detector and has energy resolving capability. Imperfect detector pixels in the detector array create ring like artifacts in the reconstruction which lower image quality. To mitigate this problem, a simple strategy based on discarding the pre-identified defective detector pixels and the appropriate rows in

the system matrix was used in this work. The explicit storage of the system matrix can be expensive for some applications but for this 2D view-limited fan-beam system, the system matrix can easily be stored in memory using a sparse matrix representation.

2.2.2 Pseudo-inverse reconstruction

The first reconstruction method is linear and is based on the pseudo-inverse (PI) of the system matrix. We use this reconstruction as a surrogate for reconstruction schemes like ART / SART.³ The explicit evaluation of the pseudo-inverse can be computationally challenging in which case a least squares approach can be used. In this case, the data model can be considered as a system of linear equations and the least-squares estimate of the image x can be obtained as:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|y - Ax\|_2^2. \quad (2)$$

This reconstruction is usually suitable for systems with $M \geq N$ (sufficient angular views). However, when dealing with view limited systems $M \ll N$ and the resulting reconstructions exhibit significant streaking artifacts due to angular undersampling. All the reference images (created from 720 views) used in this work are generated using this reconstruction scheme.

2.2.3 Total-variation reconstruction

An effective strategy to mitigate the impact of streaking artifacts that arise from view-limited imaging is to enforce spatial sparsity constraints on the object x .⁴ A popular choice for the sparsifying transform is the total-variation (TV) operator which is a function of the horizontal and vertical finite difference matrices. The sparsity constrained reconstruction problem can be posed as:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|y - Ax\|_2^2 + \lambda \operatorname{TV}(x), \quad (3)$$

where λ controls the balance between the data consistency term $\|y - Ax\|_2^2$ and the sparsity term $\operatorname{TV}(x)$. In practice, we note that the TV reconstruction is a very effective method to suppress streaking artifacts but it comes at the cost of a slight loss in apparent resolution. In this work, we use a non-linear conjugate gradient scheme with corner rounding to solve this reconstruction problem.⁵

2.2.4 Non-local rank + total-variation reconstruction

To mitigate the resolution loss caused due to a purely TV-driven reconstruction, we investigate a reconstruction model based on the low rank representation of spatial image patches. Figure 2 illustrates the core idea behind the proposed reconstruction scheme which is driven by the block matching (BM) algorithm.⁶

The entire image is divided into a set of overlapping spatial patches (or blocks). In Figure 2(a), we see that for a given exemplar patch (red block) we search a local neighborhood around this patch (dotted white region) for all the patches (non-local) that are similar to the exemplar patch. Typically, a distance metric like the ℓ_2 distance is used to identify the similar patches (blue blocks) and the top K closest blocks are identified. These similar blocks and the exemplar block are arranged to form a group (Figure 2(b)). By reshaping each of these patches into column vectors, we can form a matrix representation for the group (Figure 2(c)). As the matrix is composed of columns with very similar structure, we expect this matrix to have a very low rank which reduces the representation complexity. As we operate on non-local image patches, this approach is termed non-local rank constraint (NLR).

For an exemplar spatial patch x_i , the operator $R(x_i)$ extracts similar patches in a neighborhood. The nuclear norm, $\|\cdot\|_*$, is used to enforce low rank constraints on the extracted patches. In this framework, the TV reconstruction seen in Equation 2 is updated to incorporate the NLR constraint as follows:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|y - Ax\|_2^2 + \lambda_1 \operatorname{TV}(x) + \lambda_2 \sum_i (\|R(x_i)\|_*) \quad (4)$$

This reconstruction is solved using an alternating direction method of multipliers (ADMM) approach and involves solving a least squares problem which can be solved efficiently using a conjugate gradient based approach, soft

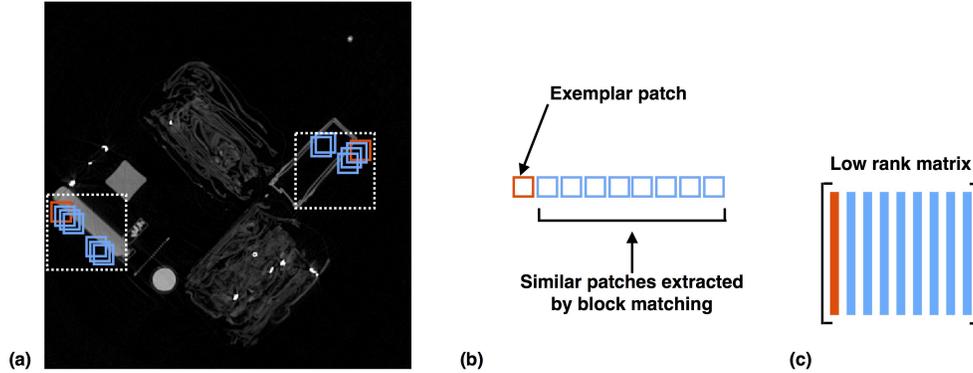


Figure 2. Illustration of block matching and non-local rank constraints. Figure (a) shows an exemplar spatial patch (red block) and closest blocks that match it (blue blocks) extracted from a local neighborhood (dotted white block), (b) shows the exemplar and extracted patches collected into a group and (c) the patches in the group are vectorized and typically exhibit a low rank structure. As the extracted patches are not restricted to the immediate neighbors of the exemplar patch, this is a non-local rank constraint. Color images can be seen in the online version of this article.

thresholding to enforce the TV constraints and singular value thresholding (SVT) for the nuclear norm.⁷ In practice, we note that this model offers a better trade-off between artifact suppression and preservation of fine details. We note that the use of the TV term is still crucial as the NLR constraint can have difficulty distinguishing undersampling artifact from image structure and can sometimes lead to overfitting artifacts.

2.3 Metrics

We use normalized mean squared error (NMSE) and high frequency error norm (HFEN) to generate quantitative performance measures for all the bags studied in this work. For a reference object x (containing N voxels) and an estimate \hat{x} , the NMSE is defined as:

$$\text{NMSE}(x, \hat{x}) = \frac{\|x - \hat{x}\|_2^2}{\frac{1}{N} \sum_i x_i \sum_i \hat{x}_i} \quad (5)$$

HFEN is obtained by Laplacian of Gaussian (LoG) filtering of x and \hat{x} and is defined as:

$$\text{HFEN}(x, \hat{x}) = \sqrt{\frac{\|\text{LoG}(x) - \text{LoG}(\hat{x})\|_F^2}{\|\text{LoG}(x)\|_F^2}} \quad (6)$$

Consistent with filter specifications in Ravishankar et al.,⁸ we use a rotationally symmetric LoG filter with a filter kernel size of 15×15 pixels and standard deviation of 1.5 pixels. Note that NMSE and HFEN are generated for each bag and we average over the bags to compare the performance of all the reconstruction algorithms on the bag ensemble. As the reconstructed images are typically processed further to create image segmentation, we used the approach proposed by Yuan et al.,⁹ to generate segmentation results for all the reconstructions. The quality of the segmentation is evaluated using the Dice coefficient (DC)¹⁰ averaged over all the segmentation labels.

3. RESULTS

In this section we present qualitative and quantitative results to compare the performances of all three reconstruction schemes. All image reconstructions are displayed using a display window set to $[0, 0.03](mm^{-1})$. For TV and NLR-TV reconstructions, the regularization parameters were optimized based on visual inspection on a few bags. These optimized values were fixed for the run on the entire bag ensemble. In addition, the NLR-TV method used 5×5 spatial patches, a patch stride length of 4 pixels, a neighborhood window of 40 pixels from which the closest 30 ($=K$) patches are extracted for each exemplar patch.

Table 1 summarizes the NMSE and HFEN results for the entire bag ensemble for the single energy system. The metrics were calculated for each bag and the mean values for each system and reconstruction method are reported. For both metrics, smaller values are better. Note that the NLR-TV reconstruction performs the best across the board with its relative performance improving as the number of views are reduced: For 23 views, the NLR-TV method is significantly better than TV whereas at 90 views, the performances of TV and NLR-TV reconstructions are very similar. The trends are similar for the dual energy system so they are not reported here.

Table 1. Averaged (over 100 bags) NMSE and HFEN results for a single energy system

NMSE (Avg)	PI	TV	NLR-TV	HFEN (Avg)	PI	TV	NLR-TV
90 views	10.83	0.74	0.71	90 views	0.82	0.38	0.35
45 views	14.97	1.13	0.91	45 views	0.95	0.49	0.45
23 views	11.49	2.04	1.28	23 views	0.99	0.63	0.56

Table 2 summarizes the quantitative comparison of the segmentations generated for all the reconstruction methods averaged across the bag ensemble. The segmentation algorithm was run both on the reference images and the reconstructions and the DC is averaged across the labels and the bags. For this metric, larger values are better as it signifies better overlap between the segmentation labels. Note that consistent with the reconstruction performance, the quality improves from PI to TV and TV to NLR-TV. Note the significant improvement that NLR-TV brings over regular TV at the 23 view case. The trends are similar in the case of the dual energy system and are not reported here.

Table 2. Dice coefficient (DC) averaged over 100 bags and segmentation labels for a single energy system

DC (Avg)	PI	TV	NLR-TV
90 views	0.408	0.758	0.803
45 views	0.391	0.639	0.765
23 views	0.352	0.517	0.685

Figure 3 shows the reconstructions for all three methods at 90, 45 and 23 view angles for a sample bag acquired with a single energy bin. Note that the PI reconstruction is significantly corrupted by the streaking artifacts. Both TV and NLR-TV reconstructions are able to reduce the streaking artifacts. Note that the NLR-TV reconstruction does a better job at retaining the fine details (red arrow) when compared with the TV reconstruction.

Figure 4 shows the reconstructions of all the methods at 45 views for both the single and dual energy systems. Note that the high energy bin of the dual energy system has significantly lower metal artifacts. The NLR-TV reconstruction again performs a good job balancing artifact suppression and recovering high frequency details (see red arrow).

Figure 5 shows the reference images, view-limited reconstructions (45 views) and the segmentations generated from both single and dual energy system. Figure 5(a) shows the reconstructions for the single energy system and the high energy bin of a dual energy system. The corresponding NMSE and HFEN results are shown at the top of each reconstruction. Note the significant improvement (both qualitatively and quantitatively) in performance when switching from the PI to TV reconstruction and the further boost attained with the use of NLR-TV reconstruction.

Figure 5(b) shows the segmentations for the single and dual energy systems. The DC scores are shown at the top of the segmentations. A higher DC score is better for segmentation quality. Note that the segmentation generated from the NLR-TV reconstruction yields the best DC score and also qualitative agrees well with the reference segmentation. Note that the segmentation for the dual energy data makes use of both the energy channels (only the high energy bin is shown in the reconstruction in Figure 5(a)).

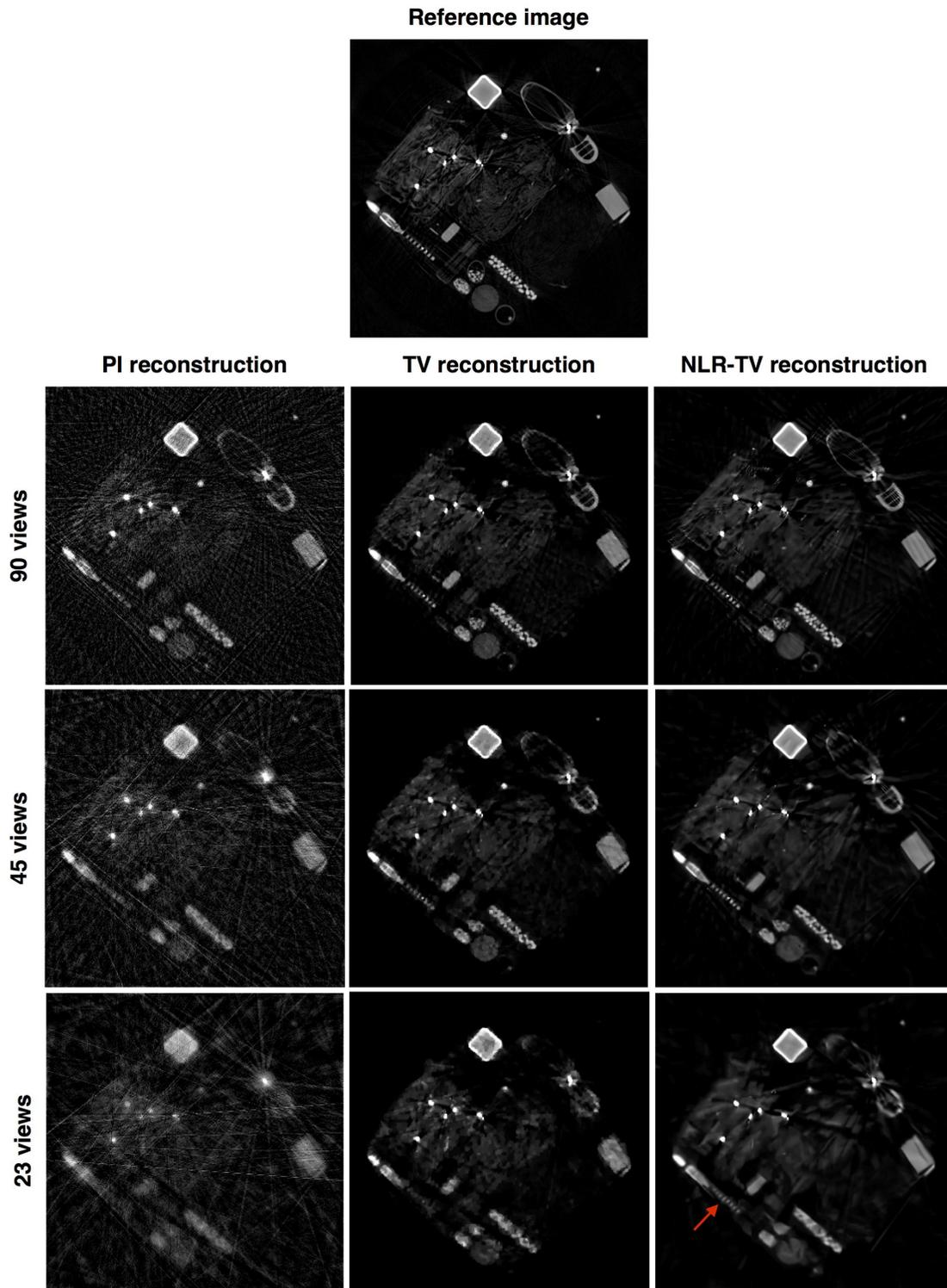


Figure 3. Image reconstructions for a single energy system at different view angles. The reference image generated from 720 views is shown in the first row. The pseudo-inverse (PI) reconstruction, total-variation (TV) reconstruction and non-local rank + TV (NLR-TV) reconstruction for 90, 45 and 23 views are shown. Note the significantly better preserved high frequency detail in the NLR-TV reconstructions (pointed by the red arrow).

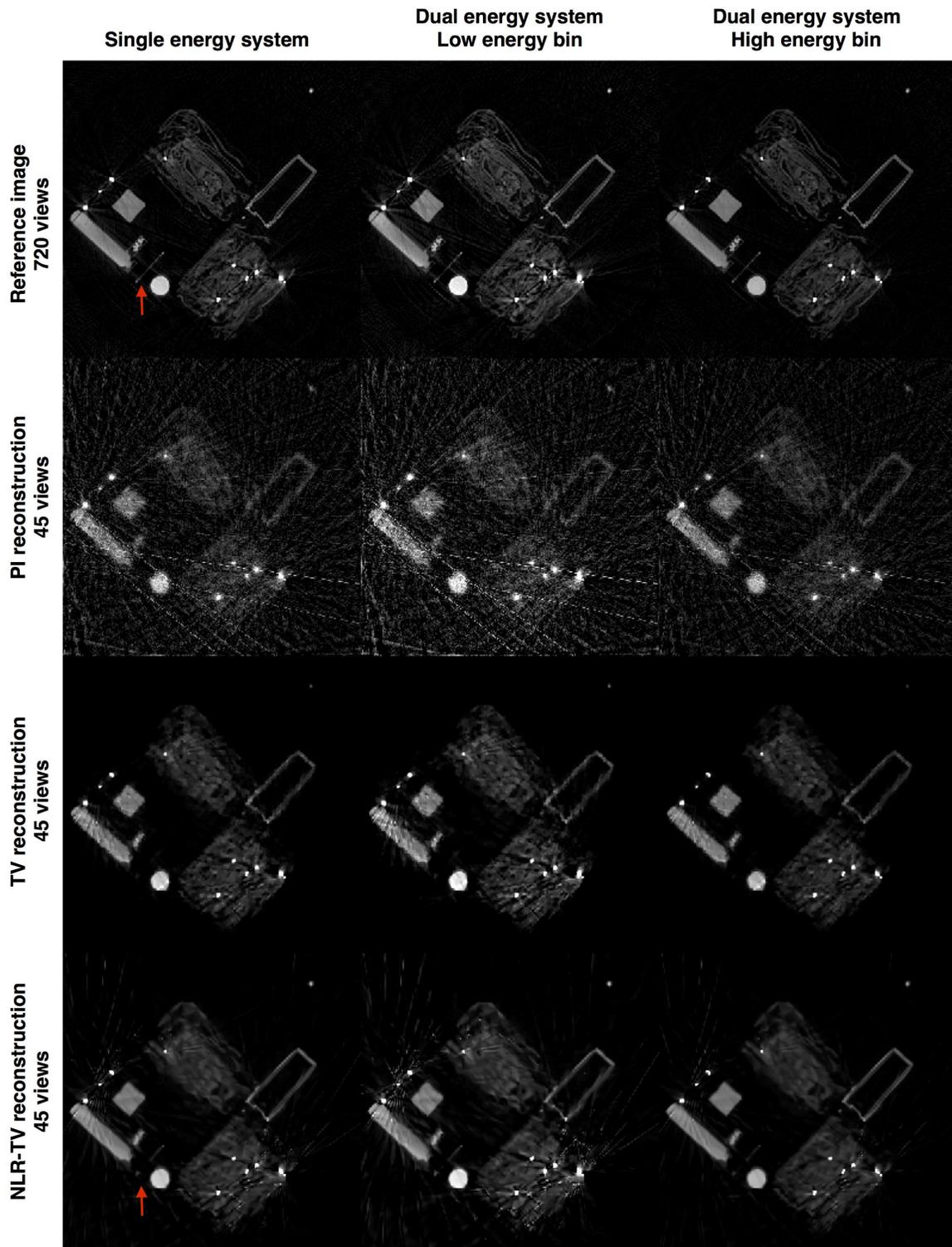


Figure 4. Comparing reconstructions across single and dual energy systems. The first row shows the reference images generated from 720 views. The subsequent rows show the PI, TV and NLR-TV reconstructions at 45 angular views. Note also the thin strip highlighted in red on the reference image is recovered only in the NLR-TV reconstruction.

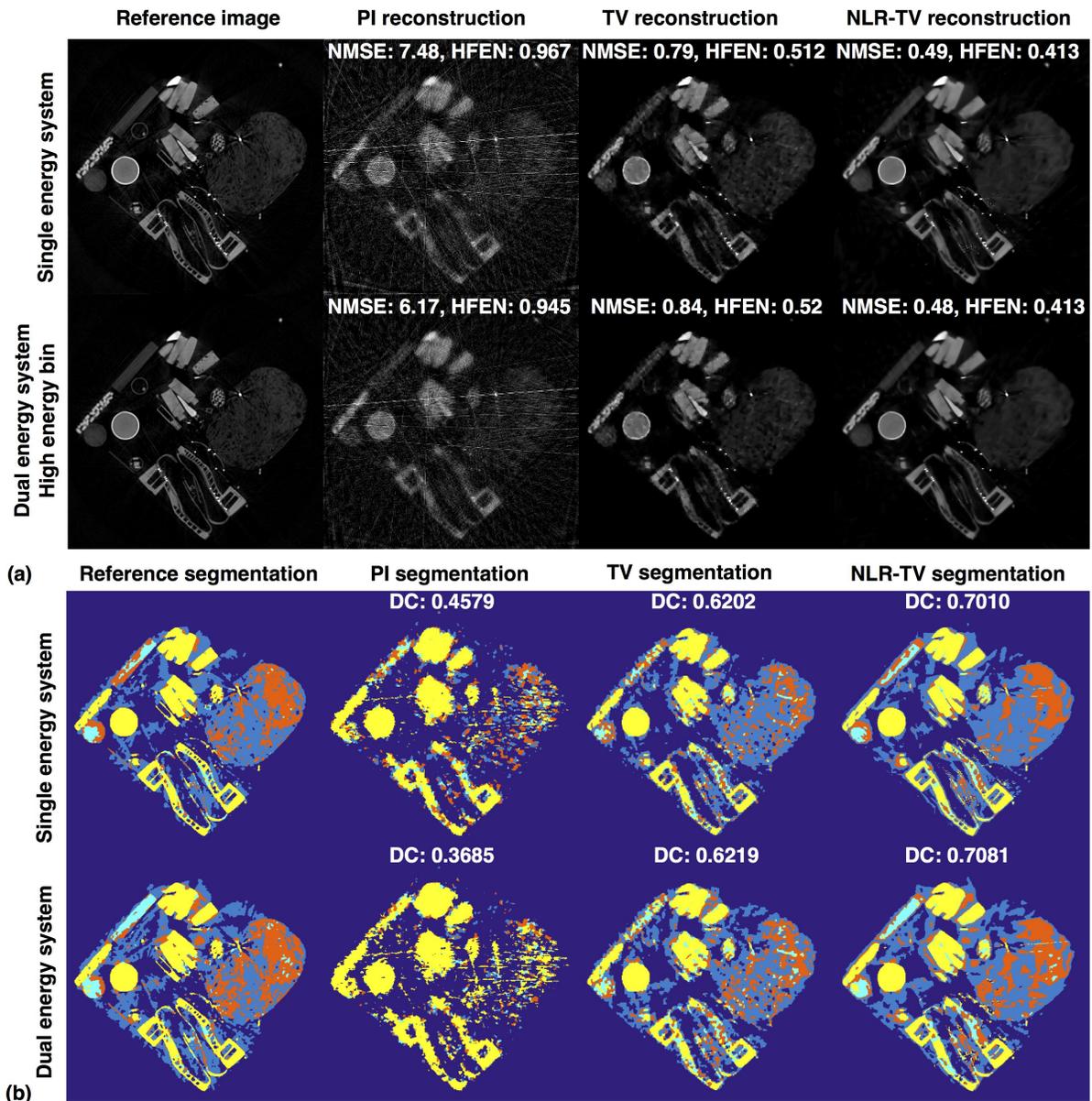


Figure 5. Reconstructions and segmentations for a sample bag. (a) Reference images (generated from 720 view data) and view-limited reconstructions (generated from 45 view data). The NMSE and HFEN results are shown at the top of each reconstruction. (b) The segmentation results for the reconstruction are shown along with the corresponding DC scores. Color images can be seen in the online version of this article.

4. CONCLUSIONS

In this paper, we studied the effects of several reconstruction schemes on realistic baggage data acquired on view-limited CT systems. We compare the PI reconstruction (surrogate for ART) and the TV reconstruction to a new reconstruction scheme, NLR-TV. NLR-TV uses the spatial sparsity constraints used by TV reconstructions and leverages an additional constraint based on the low rank structure of grouped spatial patches. We demonstrated both qualitatively and quantitatively that the constrained reconstructions (TV and NLR-TV) have significant benefits in both reconstructing and segmenting baggage data. NLR-TV reconstruction is particularly promising as it can offer a better balance between artifact suppression and preserving high resolution detail even at high view-reduction rates.

ACKNOWLEDGMENTS

This work was supported in part by the U.S. Department of Homeland Security (Department of Homeland Security) (HSHQDC-14-C-BOOIO) and the University of Arizona Technology Research Initiative Fund (TRIF) fellowship.

REFERENCES

- [1] Coccarelli, D., Greenberg, J. A., Mandava, S., Gong, Q., Huang, L. C., Ashok, A., and Gehm, M. E., "Creating an experimental testbed for information-theoretic analysis of architectures for X-ray anomaly detection," *Proc. of SPIE Conf. on Defense and Commercial Sensing*, 10187–8 (2017).
- [2] Fessler, J., "Image reconstruction toolbox (IRT)," (June 2016).
- [3] Andersen, A. H. and Kak, A. C., "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," *Ultrasonic Imaging* **6**(1), 81–94 (1984).
- [4] Bouman, C. and Sauer, K., "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. on Image Processing* **2**(3), 296–310 (1993).
- [5] Lustig, M., Donoho, D., and Pauly, J. M., "Sparse MRI: The application of compressed sensing for rapid MR imaging," *MRM* **58**(6), 1182–1195 (2007).
- [6] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K., "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. on Image Processing* **16**(8), 2080–2095 (2007).
- [7] Dong, W., Shi, G., Li, X., Ma, Y., and Huang, F., "Compressive sensing via nonlocal low-rank regularization," *IEEE Trans. on Image Processing* **23**(8), 3618–3632 (2014).
- [8] Ravishanker, S. and Bresler, Y., "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. on Medical Imaging* **30**(5), 1028–1041 (2011).
- [9] Yuan, J., Bae, E., Tai, X.-C., and Boykov, Y., "A continuous max-flow approach to Potts model," *Eur. Conf. Computer Vision (ECCV)*, 379–392 (2010).
- [10] Dice, L. R., "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).