Response to Reviewers and Associate Editor

This paper was submitted earlier, and out of the three reviewers, two of them had pronounced this as "publish unaltered" and one reviewer (R1) did not recommend resubmission. The handling Editor has encouraged a resubmission of this paper taking into account the comments of R1. We thank all reviewers for their remarks and in this document, we explain the steps we have taken to address the reviews received in our previous submission to TCI. R1 has these concerns and we quote these

> The authors have spent considerable effort in revising the manuscript and addressing the concerns of the reviewers. I would like to thank the authors for their effort. I believe the manuscript has improved because of these revisions, especially in the structure of the paper and the discussion of related work. Unfortunately, some of my main concerns with regards to the experiments, the accuracy of the proposed method, and the presentation of the findings remain present in the revised manuscript. In some cases the revisions have even increased my concerns about certain aspects. Mainly, the included experiments still do not clearly show that the proposed method improves upon existing 'simple' prior-based methods, and the new results even show that there is evidence that FDK is more accurate than the proposed method exactly in the region that changed

- **In particular, we address the following main concern of the reviewer below.**

  Note that the total variation minimization results are much better than the l1-ls results that the authors show in the paper.

  The Reviewer has taken pains in showing how the Total-Variation method is better than our earlier Compressed Sensing plus spatially-varying prior-based reconstruction. In some cases FDK is more accurate.

  This is correct for the dataset shown, and we accept this to be true.

  That said, the goal of our previous and curent work is to provide a technique that can improve upon a chosen baseline reconstruction when measurements are extremely sparse. One may choose any baseline reconstruction for pedagogical, historical, or commercial reasons, and given that, our key question is– is there a way to improve?

  Nevertheless, after noting reviewer's comments, we conducted further analysis and observed that TV is indeed better suited as a baseline reconstruction method for our datasets and our subsampling choices. Hence, our current paper presents all results using TV regularization coupled with spatially-varying technique, and demonstrates its benefits over TV-only and backprojection-only methods. We have eliminated our previous implicit claim that the CS based scheme is optimal for the dataset provided.

  As far as FDK is concerned, the current results indeed show the superiority of the TV-based scheme (as also noted by the reviewer).

- **Other comments given by Reviewer R1 and the handling editor**
  - ...the paper should at least include a (short) discussion about the possibility of using deep learning for this, and its disadvantages.

    We have now discussed the difficulties in using deep-learning techniques for the current problem in Sec. 7C of our paper. We also describe them here below.

    In a machine learning based setting, our goal of detecting new changes is a 'prediction' problem in a continuous solution space i.e., *"given intensities of a voxel at various time instants in a longitudinal setting and partial measurements of the voxel at the current time, what will be the intensity of the same voxel at the current time instant?"* This estimation can be learnt by a deep neural network if there are hundreds of labeled data. However, generalization of this across multiple datasets poses a question mark.

    Further, a specific 'event-of-interest' may occur just once in a longitudinal study and hence training on data of all previous time instants may be misleading. If however, we use measurements from identical and full-cycle longitudinal studies, including events-of-interest, of similar specimen (instead of the 'same' specimen as we used), a deep-learning technique may be applied. We see this as an extension of using object-prior generated from the same object, and hence we have not explored this direction.

    Another avenue for using deep-networks *within* our current work is to replace the currently used fixed eigenspace provided by PCA by a learnt feature basis provided by a network such as an autoencoder. For each dataset, an autoencoder may be trained to learn a specific set of latent features (this might again require atleast a few tens of volumes). We see this as a future direction of work.
  - There are no added comparisons with popular existing methods for reconstruction with a small number of projections (TV, algebraic techniques, etcetera)

    Among the family of algebraic techniques, we have compared our method with Algebraic Reconstruction Technique (ART) and Simultaneous Algebraic Reconstruction Technique (SART). We do not reconstruct using Discrete Algebraic Reconstruction Technique (DART) since there is no prior information about the number of possible attentuation coefficients in the test object. Among other iterative techniques, we have used Simultaneous Iterative Reconstruction Technique (SIRT), Total Variation (TV) and Compressed Sensing (CS) with Haar wavelet and DCT as sparsity basis. The details of solvers used for these algorithms and their reconstructions are in Page. 7 (Section-6A) of the paper. For CS reconstruction, we observed that Haar wavelet and DCT were best suited for our dataset among various other possible sparsity basis.
  - Why did the authors not choose the standard settings for SSIM?...I don't understand the reason given for not including RMSE: the values don't have to be optimized again – you could just optimize for SSIM but at least still report RMSE comparisons.

    In our experiments, we observed that for a given dataset and a set of projection measurements, the intensity span of histograms of reconstructed volumes differ across various methods and solvers. Hence, we choose to focus on preserving the structures on our reconstructions, and not rely on the absolute intensity values alone. Hence, we choose SSIM with a higher weightage to structure-preservation. For our record, we have computed RMSE values as well and they are presented here in Tables 1 and 2.

Table 1: RMSE within the RoI of 3D reconstructed Potato volume from various methods.

Backprojection	**TV**	**This paper**
0.29	0.24	0.18

Table 2: RMSE within the RoI of 3D reconstructed Sprouts volume from various methods.

Backprojection	**TV**	**This paper**
0.53	0.48	0.30

Table 3: Computation times (in seconds) for various stages of reconstruction of a 2D slice of Liver on MATLAB 2019a on an AMD 2920X 12-Core Processor machine with 64GB RAM. Here we assume that low quality eigen-spaces of the templates are computed offline.

Backprojection	TV	Computation of weights map	Final reconstruction with spatially-varying prior
0.40	8.44	0.04	46.73

– A discussion about computational costs (time, memory) should be included in the paper.

We now present details of both time and space complexities of our technique in Section.7B of our paper.

Most of the computation time in our technique is used for the generation of low quality eigen-space by reconstructing each of the templates using TV and FDK. If the set of projection views for every test is fixed a priori, the set of low quality reconstructions can be produced offline and stored in order to save computational costs. With this assumption, Table 3 shows the computation times for various stages of reconstruction of a 2D slice of Liver on MATLAB 2019a on an AMD 2920X 12-Core Processor machine with 64GB RAM. If however, new sets of projection views for every test are to be allowed, the low quality reconstructions can still be performed efficiently using parallelization since these reconstructions are completely independent. Here below, we describe the time complexity of spatially-varying prior technique. If

* N = number of voxels of the volume
* T = number of templates available, and
* F = time taken for iterative TV optimization,

then, the computational complexity is sum of the following

* $O(NT^2 + T^3)$ for creating the eigenspace,
* $O(NT^2)$ for deriving the weights map, and
* time F taken for iterative minimization

where $O(.)$ is the notation for 'Big O' used in complexity analysis. In terms of space, the complexity is $O(T)$ in order to store the templates and their low quality reconstructions. We only need to store a single weights map volume regardless of the number of templates available.