Discrete Iterative Partial Segmentation Technique (DIPS) for Tomographic Reconstruction

Toby Sanders

Discrete Iterative Partial Segmentation Technique (DIPS) for Tomographic Reconstruction

Toby Sanders

Abstract

In this paper, we propose a new reconstruction technique for discrete tomography. In general, discrete tomography refers to reconstructing images or volumes from projection by placing a strong prior that the objects to be reconstructed contain only a few intensities. Our technique includes a soft segmentation and an refinement update in each iteration, only segmenting pixels that are near the known densities of the images objects. In this way the dimension of the reconstruction problem is carefully and slowly reduced as the structures become more evident. In addition, we include additional prior knowledge in each refinement update through compressive sensing techniques, which proves to be extremely beneficial to the method. We perform simulation experiments that show substantial improvement in the accuracy of the reconstruction in several different scenarios.

1 Introduction

Tomography is a well-established technique for non-invasive 3-D imaging, with many applications within medicine, biology, and materials science [1–3]. The work flow for tomography typically includes first collecting projections of an object from several different angles and then using these projections to reconstruct a full approximation of the object. The reconstruction problem is an ill-posed inverse problem that is often underdetermined [4], and therefore many different reconstruction approaches have been developed to improve the resulting approximation.

Traditional algebraic techniques [5,6], such as the algebraic reconstruction technique (ART) and the simultaneous iterative reconstruction technique (SIRT), are designed as iterative methods to efficiently reconstruct an object that quantifiably satisfies the projection measurements, typically without any other prior knowledge. However, in many cases these simple approaches are deficient [7], such as when the data is limited and noisy, or the object has a very complex structure. This is particularly evident whenever the reconstruction containing a continuous range of densities must be nontrivially separated (or segmented) into different uniform structures.

Therefore, to resolve these issues reconstruction models should look to incorporate additional available knowledge. Most commonly, reconstruction techniques benefit from incorporating additional prior knowledge of the object through smoothness characterization of the images. In many tomographic settings one can often define this smoothness through the prior knowledge that the scanned objects are known to have only a few compositions (discrete structures in the segmentation), so that only a few intensities (or gray values) may be needed to represent the structures. In this paper we discuss two types of regularization techniques that incorporate this knowledge.

The first is a more general and less restrictive approach known as total variation (TV) minimization [8], a regularization model that encourages a sparse boundary set in the reconstructed images. This approach has proven as a practical method for image reconstruction in many fields over the last several years, including tomography [9–11]. While TV minimization can significantly improve quality of the solution, almost surely the model will not recover a completely discrete solution. Therefore error may still be evident in the segmentation of images from this approach.

The second is discrete tomography, which deals with the study of tomographic reconstruction of images with only a few prior selected gray values [12, 13]. Many techniques within discrete tomography deal with very special sampling cases and reconstruction of very special objects, such as reconstructing binary images with only a few projections [14–16]. However, perhaps a more practical discrete tomography method is what is known as the discrete algebraic reconstruction technique [17] (DART), which has been used to reconstruct carbon nanotubes, zeolites, and other nanoparticles [18–20]. This technique aims to reconstruction an object that quantifiably satisfies the projection measurements with only several prior selected gray values, by applying a rigid threshold to an initial solution and iteratively correcting this solution with an algebraic reconstruction method. While such a method has the benefit of recovering a solution that is completely discrete, the rigid nature of these procedure lends towards convergence of some local minimum near the initial threshold as opposed to an optimal global minimum solution.

Several alternatives to DART have since been recently suggested. For instance, an adaptive DART [21] was shown to improve the approximation of very fine local structures at the boundary. A softer version of DART known as SDART [22] was implemented to improve the model for noisy data. However, these approaches do not generally handle the issue of convergence to a global minimum.

In this paper, we present a new discrete tomographic reconstruction technique, which implements a modified TV regularization technique based on partial segmentations of the reconstruction. As in discrete tomography, we must assume that the objects are truly homogenous and choose a few target densities. After an initial solution is computed, the volume is partially segmented, by classifying only pixels that are very near to one of the target densities. The remaining pixels are consider uncertain, however based on new knowledge of pixels that were classified we refine the solution by minimizing an appropriate TV regularization model exploiting sparsity of the images, and this procedure it iterated until some convergence is met. In this way, the dimension of the reconstruction problem is slowly reduced as the segmentation classes become more clear, which ideally will avoid arriving at a very inaccurate segmentation, leaving greater possibility for convergence to a true global minimum.

The outline of the paper is the following. In section 2 we first define some notation and clearly define the tomographic reconstruction problem and projection model. We then give details of the existing methods TV minimization and DART in sections 3 and 4. Our new reconstruction approach is clearly defined and outlined in section 5. Finally, in section 6 we give numerical results, comparing the methods with several different phantom images of various complexity and various projection quality. The reconstruction techniques are first compared whenever the projections are limited and/or noisy. We then give results whenever the angular tilt at which the objects can be projected is limited, a common problem in electron tomography known at the "missing-wedge."

2 Notation and Tomographic Reconstruction Problem

Let f be a vector in $\mathbb{R}^n$ and W be a matrix in $\mathbb{R}^{m \times n}$. We denote the ith entry of f by f_i and the (i, j) entry of W by $w_{i,j}$. If we let $T \subset \{1, 2, \ldots, n\}$, then the vector f_T denotes the vector in $\mathbb{R}^{|T|}$ restricted to only the entries of f from the set T. Similarly, we write W_T to be the matrix in $\mathbb{R}^{m \times |T|}$ restricted to the columns of W from T.

For $p \geq 1$, the ℓ_p norm of f is defined by

$$\|f\|_p = \left(\sum_{i=1}^n |f_i|^p \right)^{1/p}. \tag{1}$$

Images will be considered pixelated $n \times n$ matrices or vectors. For most computational purposes an image f will be given a vector, $f = \{f_i\}_{i=1}^{n^2}$. In other instances f may be considered as a matrix, $f = \{f_{i,j}\}_{i,j=1}^n$. Each representation will be considered an equivalent re-indexing of the other.
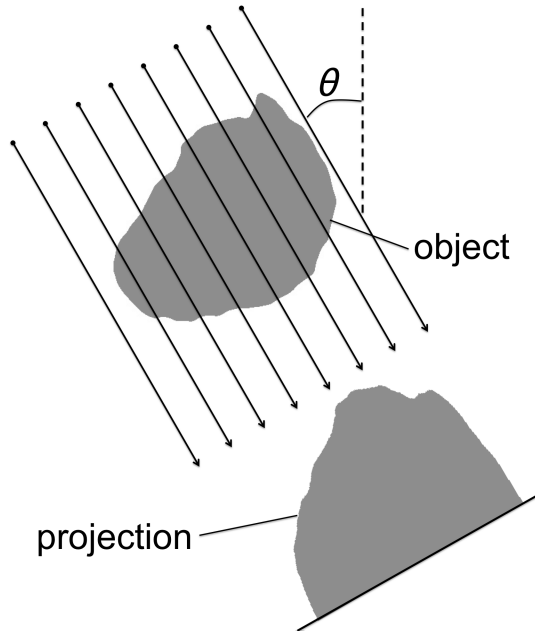
Figure 1: A parallel set of rays are used to project the object at an angle of $\theta = 30°$.

For this paper, we consider the tomographic reconstruction problem for only 2-D images, although the procedure is generalized to higher dimensions in a straight forward manner. A projection at angle θ of a 2-D image can be defined to be the set of all line integrals of the image at angle θ with respect to some coordinate system. Shown in figure 1 is a single projection of a simple 2-D object taken at an angle $\theta = 30°$. Formally, one may define the projection of a function or image f on a continuous domain through the Radon transform at angle θ, $R_\theta : \mathbb{R}^2 \to \mathbb{R}$, defined by

$$R_\theta f(x) = \int_{\mathbb{R}} f((x\ y)Q_\theta)\ dy, \tag{2}$$

where

$$Q_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

These projections are acquired at various angles θ and are use as data to reconstruct an approximation of the object. A reconstruction from these projections is a pixelated $n \times n$ image f. In practice, a single value of a projection representing the integral about some linear path L will be recorded by a detector to be some value, say b_i. In terms of our image, this integral may approximated through our $n \times n$ discretization of f analogous to a Riemann sum, so that

$$b_i = \sum_{j=1}^{n^2} w_{i,j} f_j, \tag{3}$$

where $w_{i,j}$ is the length or area of the intersection of the path of L with pixel f_j. Given a total of m such projection values $b_i,\ i = 1, 2, \ldots, m$, we write equation (3) for each b_i and obtain the linear system

$$Wf = b, \tag{4}$$

where $W = \{w_{i,j}\}_{i=1,j=1}^{m,n^2}$ and $b = \{b_i\}_{i=1}^{m}$. Given this system, most reconstruction techniques rely on finding a solution f nearly satisfying (4), so that the reconstruction quantitatively agrees

3

with the projections. In many tomography settings, the amount of data is much smaller than the needed resolution of the reconstruction image, leaving $m \ll n^2$. This leads to a severely underdetermined problem, so that there may theoretically exist infinitely many unique solutions nearly satisfying (4), which further points for the need to search for special solutions with desired smoothness properties.

3 Total Variation Minimization

Total variation minimization is a special case of compressed sensing, which is in general is a powerful mathematical model for recovery of sparse signals whenever the information of the unknown signal is severely limited [23–25]. Given an unknown sparse signal, say f^*, classical theory of compressed sensing tells us that under special sampling conditions, if f^* is sparse enough then it will be recovered by finding the solution that minimizes $\|f\|_1$. In general, the unknown signal, say f^*, may not be sparse, but sparse perhaps under some transformation, i.e. there exist some operator U so that Uf^* is sparse. For tomographic reconstruction of discrete objects, we assume that the boundary of the true objects are sparse. Therefore an appropriate transformation of the object to give sparsity is the 2-D discrete gradient, so that perhaps ∇f^* is sparse. Therefore, the reconstruction method of TV minimization relies on finding a solution minimizing $\|\nabla f\|_1$, known as the total variation norm, denoted $\|f\|_{TV}$. For a precise definition, we first re-index f with double indices, $f = \{f_{i,j}\}_{i,j=1}^n$, and then the TV norm is defined by

$$\|f\|_{TV} = \sum_{i=1}^{n} \sum_{j=1}^{n-1} |f_{i,j+1} - f_{i,j}| + \sum_{j=1}^{n} \sum_{i=1}^{n-1} |f_{i+1,j} - f_{i,j}|. \tag{5}$$

Finally, to account for noise the general model does not require a solution exactly satisfying (4) and allows the solution to deviate slightly from the recorded projections. Therefore the usual model for TV minimization becomes to find the solution to

$$\min_{f \in \mathbb{R}^{n^2}} \left\{ \|f\|_{TV} + \lambda \|Wf - b\|_2^2 \right\}. \tag{6}$$

Larger selection of λ will encourage the solution to more accurately satisfy the data. Smaller values of λ will enforce greater regularization, and should be used whenever more noise is evident. For our work, in order to solve (6) we use the TVAL3 scheme [26], which is based on minimization of augmented Lagrangian functions through alternating direction minimization scheme. The algorithm is implemented in Matlab by *Li et. al* [26].

4 Discrete Algebraic Reconstruction Technique

DART [17] takes a much tighter grip on enforcing the fact that there should ideally only be a few compositions within the reconstructed objects. Prior to reconstruction, gray value are estimated and fixed for each material composition, as well as corresponding thresholds for the segmentation steps. Let us denote the gray values by $\boldsymbol{\rho} = \{\rho_1, \rho_2, \ldots, \rho_k\}$ and the thresholds by $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_{k-1}\}$, where $\rho_1 < \tau_1 < \rho_2 < \tau_2 < \cdots < \tau_{k-1} < \rho_k$. To our knowledge, DART was designed as a practical method to find the best solution that contains only the prior selected gray values, which can be expressed as finding

$$\min_{f \in \mathbb{R}^{n^2}} \|Wf - b\|_2 \quad \text{s.t.} \quad f_i \in \boldsymbol{\rho} \tag{7}$$

for $i = 1, 2, \ldots, n^2$. In order to approximate such a solution, DART combines a segmentation with an update using continuous algebraic reconstruction method. In each iteration of DART the

solution is fully segmented and this segmented solution is refined along the boundary with the continuous algebraic reconstruction method, eventually converging to a solution which hopefully realizes something close to (7). We outline the method with the following steps:

1. An initial reconstruction f^0 is computed using a algebraic method, such as SIRT.

2. The current reconstruction is segmented using thresholding, based on $\boldsymbol{\tau}$ and $\boldsymbol{\rho}$.

3. The boundary pixels (pixels that have at least one neighbor with a different gray value in the 8 pixel neighborhood) of the current segmented solution are set as free. Additional randomly selected pixels may be set to free as well. The remaining pixels are fixed, and their contribution to the projections are subtracted from the projection measurements.

4. An update solution is computed by solving the reduced system involving only the free pixels by applying an algebraic reconstruction method.

5. Smooth the solution by convolution with kernel. If convergence is not met, return to step 2.

The segmentation in step 2 is done using basic thresholding, which can be represented by a thresholding function, $T_{\boldsymbol{\rho},\boldsymbol{\tau}}$, which we will sometimes write as T. For an image f, $T_{\boldsymbol{\rho},\boldsymbol{\tau}}$ maps f to a new thresholded vector, where the ith entry is given by

$$T_{\boldsymbol{\rho},\boldsymbol{\tau}}(f_i) = \begin{cases} \rho_1 & \text{if } f_i < \tau_1 \\ \rho_2 & \text{if } \tau_1 \leq f_i < \tau_2 \\ \vdots \\ \rho_{k-1} & \text{if } \tau_{k-2} \leq f_i < \tau_{k-1} \\ \rho_k & \text{if } \tau_{k-1} \leq f_i \end{cases}. \tag{8}$$

While the algebraic reconstruction method used to compute the initial solution was designed to quantitatively satisfy the projection data, the segmentation of this solution may not, so that

$$\|WT(f^0) - b\|_2 \gg \|Wf^0 - b\|_2. \tag{9}$$

Therefore the solution is corrected by refinement with an algebraic reconstruction method along the boundary of the segmented image. At iteration t, given the current solution f^t and the segmentation of the solution $T(f^t)$, let us denote the set of free pixels (boundary pixels in $T(f^t)$) by $B \subset \{1, 2, \ldots, n^2\}$ and the set of fixed pixels by $F = B^c$. In addition, some randomly selected pixels from the fixed set F may set as free pixels, with probability p. Removing the contribution of the fixed pixels from the measurements, we obtain new measurements $b_B = b - W_F T(f^t)_F$, which ideally would represent the projection data for only the free pixels B. Then the new updated solution along the free pixels is given as

$$f_B^{t+1} = \underset{f_B \in \mathbb{R}^{|B|}}{\operatorname{argmin}} \|W_B f_B - b_B\|_2, \tag{10}$$

which can be solved using an algebraic reconstruction method. The updated solution remains the relatively unchanged over the fixed pixels, although the pixels near a free pixel can change in step 5 when the solution is then smoothed by convolution with a kernel.

Convergence can be given in several forms. As mentioned in other papers, one can either set a fixed number of iterations or set an appropriate ε tolerance on the change in the solution so that it converges when $\|f^{t+1} - f^t\|_2 < \varepsilon$.

5   Discrete Iterative Partial Segmentation Technique (DIPS)

As mentioned before, DART was designed as a practical method in hopes of finding the discrete solution which most accurately satisfies the projection data, i.e. to solve (7). However, it's possible the initial segmented solution may have misclassified a large number of pixels, which may be clear from a large jump in the projection error mentioned in (9). It is not completely clear how well this refinement of the solution will correct this misclassification. In our experience, we have found that DART is highly subjugated to the initial starting solution and initial segmentation of this solution, converging to a local minimum near this point.

In the following section we outline a new method that involves partial segmentation, only classifying those pixels for which we have the greatest certainty. We then refine the solution instead over the regions which were not classified, and only the classified pixels are fixed. In this way, we soften the segmentation process, and gradually reduce the dimension of the problem as the classifications become more clear. In addition to this, in section 5.2 we provide a crucial improvement in the minimization procedures, providing a practical way to use TV regularization for the update steps.

5.1   Partial Segmentation and Region Refinements

Here we offer a simple strategy to avoid jumping to a poorly segmented solution through a sequence of softened partial segmentations and refinements. Similar to DART, we will compute an initial solution and iteratively update this solution. However, we define segmentation and refinement operations that are significantly relaxed to avoid steep degradation of the projection error. Let our gray values be given by $\boldsymbol{\rho} = \{\rho_1, \rho_2, \ldots, \rho_k\}$, with $\rho_1 < \rho_2 < \cdots < \rho_k$. Alternative to rigid threshold values, we define segmentation tolerances or radii $\mathbf{r} = \{r_\ell\}_{\ell=1}^k$, and define the balls around each interior gray value

$$A_\ell(r_\ell) = (\rho_\ell - r_\ell, \rho_\ell + r_\ell), \tag{11}$$

for $\ell = 2, 3, \ldots, k-1$, and the sets

$$A_1(r_1) = (-\infty, \rho_1 + r_1),$$
$$A_k(r_k) = (\rho_k - r_k, \infty).$$

Now, in each iteration only the pixels that are within these neighborhoods of the target gray values are segmented. In particular, pixels are only segmented if they are in the set $\cup_{\ell=1}^k A_\ell(r_\ell)$. Our soft segmentation function $S_{\boldsymbol{\rho}}$ will map f to a partially segmented solution, where the ith entry is

$$S_{\boldsymbol{\rho}}(f_i; \mathbf{r}) = \begin{cases} \rho_1 & \text{if } f_i \in A_1(r_1) \\ \rho_2 & \text{if } f_i \in A_2(r_2) \\ \vdots \\ \rho_k & \text{if } f_i \in A_k(r_k) \\ f_i & \text{if } f_i \notin \cup_{\ell=1}^k A_\ell(r_\ell) \end{cases} \tag{12}$$

We note that in order for $S_{\boldsymbol{\rho}}(\cdot\,; \mathbf{r})$ to be well defined, our sets should not overlap, so that the tolerances $\{r_\ell\}_{\ell=1}^k$ should be chosen so that

$$A_{\ell_1} \cap A_{\ell_2} = \varnothing \tag{13}$$

for

$$\ell_1 \neq \ell_2, \, \ell_1, \ell_2 = 1, 2, \ldots, k.$$

With this softened segmentation, we have only classified the pixels that are near the target gray values and have the greater probability that they are being correctly classified. The solution is then refined on a reduced system containing the unclassified pixels. The free set is defined by

$$R = \left\{ i \mid f_i \notin \bigcup_{\ell=1}^{k} A_\ell(r_\ell) \right\} \tag{14}$$

and the fixed set is given by

$$F = R^c = \left\{ i \mid f_i \in \bigcup_{\ell=1}^{k} A_\ell(r_\ell) \right\}. \tag{15}$$

As a point of emphasis, we denote the free set with R (for region), because the free pixels are not randomly scattered about the image but rather bundled into regions on the image domain. In addition the values of the free pixels are bundled into regions on the real number line lying between the gray values. As with DART, we may additionally choose to randomly assign some pixels from the fixed set to be free pixels, with probability p. The additional steps in the method remain relatively consistent with the traditional DART. In particular, at some iteration t, given the free pixel set R and fixed pixel set F, the updated projection data representing only R is given as $b_R = b - W_F S_{\boldsymbol{\rho}}(f^t; \mathbf{r})_F$, and the updated solution over R is then given by

$$f_R^{t+1} = \operatorname*{argmin}_{f_R \in \mathbb{R}^{|R|}} \|W_R f_R - b_R\|_2, \tag{16}$$

which is the analog of (10). For regularization purposes, the updated solution is then smoothed by convolution with a kernel. While the general steps are the same, making these changes significantly modifies the algorithm and can have significant impact on the results. Let us clearly note the two important points in the method:

1. The segmenting function is softened by using $S_{\boldsymbol{\rho}}(\cdot, \mathbf{r})$ instead of $T_{\boldsymbol{\rho},\boldsymbol{\tau}}(\cdot)$ for partial segmentation in order to more carefully classify the pixels.

2. The free pixels are defined by R instead of the boundary pixels B, in order to account for the segmentation function and redefine the pixels used for refinement.

Finally, we mention that this procedure can simply serve as a precursor to using the traditional DART, in order for the DART iterations to begin with a fully segmented solution that is perhaps closer to a true global minimum. Once there remain very few pixels in R and the solution appears to have converged, we proceed by computing full segmentations with $T_{\boldsymbol{\rho},\boldsymbol{\tau}}$ (for some chosen thresholds $\boldsymbol{\tau}$) and setting the the free pixels as the boundary pixels, B. We will refer to the technique described in this section as DIPS-LS (least squares), to indicate that the minimization in (16) in each iteration is the least squares minimization problem. We now suggest a way to improve this minimization model.

5.2 Implementation of Region Refinements with TV Minimization

In this section, we offer a crucial technique for improving our DIP-LS approach in the refinement step by making use of TV minimization by solving a variation of (6) for the update, instead of solving (16). We refer to this approach simply as DIPS.

As discussed earlier, TV minimization is a method for recovering solutions that are assumed to have sparse boundary regions. With our iterative partial segmentation, for the first several iterations the set of free pixels may be very large, and they are likely bundled into large connected regions over the image as opposed to being randomly scattered about the image. This holds especially true if TV minimization was used to compute these solutions and encourage this type of

smoothness. Therefore we may still assume the objects being reconstructed have a sparse boundary region over R, and it is reasonable to refine these regions with a minimization that includes the TV norm from (5) to improve the model. We emphasize that inclusion of the TV norm is only useful given relatively large regions of free pixels and is likely not beneficial for implementation with the traditional DART, where the free pixels are already the sparse boundary regions, B.

This method will be implemented precisely as DIPS-LS, with the only change given in the minimization procedure from (10). As before, the solution at iteration t is f^t, with the free pixel set R and fixed pixel set F. Then we again take the updated projections to be $b_R = b - W_F S_{\boldsymbol{\rho}}(f^t; \mathbf{r})_F$, and the updated solution is now given by

$$f^{t+1} = \underset{f \in \mathbb{R}^{n^2}}{\operatorname{argmin}} \left\{ \|f\|_{TV} + \lambda_1 \|Wf_R - b_R\|_2^2 + \lambda_2 \|f_F - f_F^t\|_2^2 \right\}. \tag{17}$$

As we will show, minimization of this functional proves to be a very beneficial alternative to (16). The TV norm in the first term of (17) is again to encourage appropriate smoothness of the solution and sharp boundaries within R. Ideally, one would only need some special adjusted TV regularization norm to serve only on R, but for simplicity we simply include the TV criteria over the entire solution. Also, the TV norm over the segmented set is already very small, and therefore change the values over F would unlikely result in a smaller minimum. However, to further ensure that the solution does not deviate over F, the last term in (17) penalizes such occurrences. In our practice, we set $\lambda_2 = 10 \cdot \lambda_1$, which was found simply through experimentation and can be increased further if changes in the fixed pixels is observed. The middle ℓ_2 term is consistent with the minimization for refinement used in (10) and (16). Minimizing (17) requires significantly more computational time than solving (16). Additionally, it is only practical if R is relatively large and connected. Therefore, we minimize (17) for only a few iterations, typically about 10 to 15, or until R is relatively sparse. After these iterations, we proceed as before by "cleaning" the boundaries using the traditional DART. We now will briefly discuss the appropriate choice of the tolerance values, $\mathbf{r} = \{r_\ell\}_{\ell=1}^k$.

5.3   Choice of the Segmentation Tolerance

The chosen tolerance values will have an effect to both the accuracy of the reconstruction and the rate of convergence. For larger tolerance values, the pixels will fall into their classifications more quickly and speed-up the convergence of the algorithm. On the other hand, smaller tolerance values will make the segmentation sets smaller so that fewer pixels will fall into these set keeping the dimension of the problem large, slowing the rate of convergence. However, it may be worthwhile to do so, in order to increase the likelihood that these pixels, when classified, are put into the correct classification. In addition, due to refinement procedure in (16), those pixels which are misclassified in F not only directly increase deviation from the true solution but will further induce error onto the refinement over R.

We suggest to allow the tolerances to vary within each iteration using a greedy type of approach. To begin, start with a very small, rigid tolerance and only pixels which are very near to the target gray values are segmented. As the iterations progress, if very few new pixels are being classified, so that R and F are not changing and R is not tending towards zero, then increase the tolerance slightly and proceed. This can be expressed in the following way. At some iteration t, let R^t and F^t denote the free and fixed pixel sets at this iteration and $\{r_\ell^t\}_{\ell=1}^k$ denote the segmentation tolerances. If R^t and R^{t-1} are relatively equal, then we increase each tolerance by some small δ in the next iteration so that $r_\ell^{t+1} = r_\ell^t + \delta$, for $\ell = 1, 2, \ldots, k$. We can say that two sets are relatively equal if the cardinality of their intersection is relatively equivalent to the cardinality of their union. For example, the sets R^t and R^{t-1} are relatively equivalent with tolerance ε if

$$\frac{|R^t \cup R^{t-1}| - |R^t \cap R^{t-1}|}{|R^t \cup R^{t-1}|} < \varepsilon \tag{18}$$

8

The region refinements can then be said to converge whenever $|R^t| \approx 0$ or the unfortunate circumstance that two of the tolerances r_ℓ^t and $r_{\ell+1}^t$ are large enough so that the segmentation sets $A_\ell(r_\ell^t)$ and $A_{\ell+1}(r_{\ell+1}^t)$ overlap. For all practical purposes, $|R^t|$ never actually converges to zero, because the smoothing step results in boundary pixels moving out of the segmentation classes, and so at best $|R^t|$ actually converges to a "boundary-like" set, in which case the method practically converges to something similar to DART.
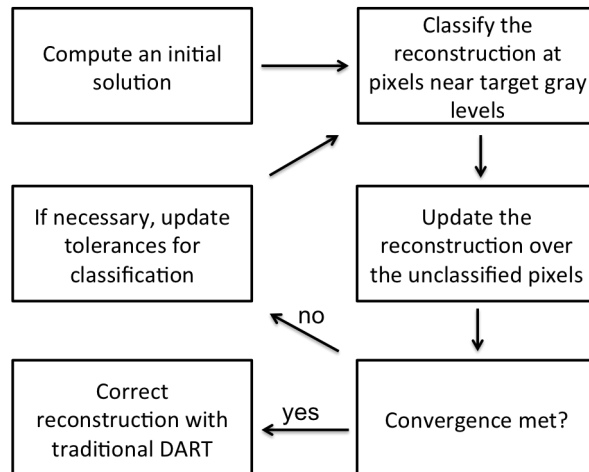
Figure 2: General procedure for DIPS.

6   Results

To compare these reconstruction approaches for discrete tomography, we perform simulations and analysis on a set of five phantom images. The phantom images that are used are shown in Figure-3. Each of these images are contained in a 512×512 pixel grid. Phantoms 1, 4, and 5 are all binary with gray values $\boldsymbol{\rho} = \{0, 1\}$. Phantom 2 and 3 have gray levels $\boldsymbol{\rho} = \{0, 0.5, 1\}$. Phantom 5 was taken from the original DART paper [17] for a base comparison with the original results.

In each simulation, we compare the performance of each of the described discrete reconstruction methods, DART, DIPS-LS, and DIPS. For a base comparison with a traditional algebraic reconstruction method we also provide results from reconstructions with SIRT. In order to quantify the error of approximation, the relative number of misclassified pixels is calculated, which we denote rNMP. We formally define the rNMP as the fraction of pixels in the resulting reconstruction that
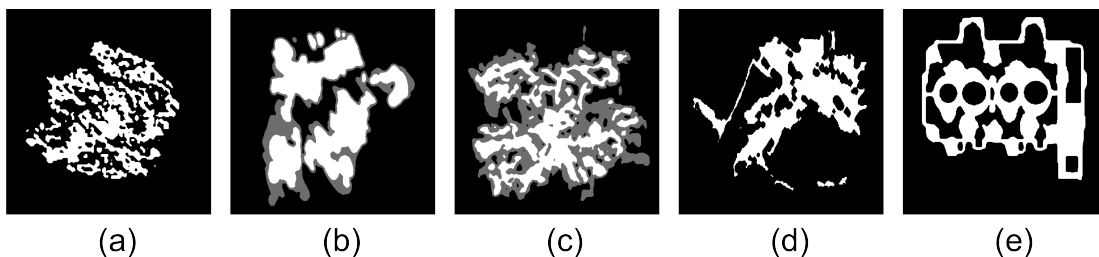

Figure 3: Shown are the phantom images used for simulations. The phantom images are referred to as phantom 1 (a), phantom 2 (b), phantom3 (c), phantom 4 (d), and phantom 5 (e).

9

do not agree with the original phantom.

For the SIRT reconstruction, we perform a total of 200 iterations. A density constraint is implemented, where at the end of each iteration the negative pixels are set to the minimum density, zero. This constraint only improves the results, so that we report only results with the density constraint in this paper. In order to segment the SIRT reconstruction, we do not use any technical segmentation techniques such as Otsu segmentation, as we actually find that these methods only increase the rNMP. Instead, we simply use the thresholding function T, where the threshold values are set to be the averages of the gray values.

A total of 200 iterations are performed with DART. The initial solution is taken from the SIRT reconstruction mentioned above. The algebraic update steps to solve the intermediate problems (10) are performed using 20 iterations of SIRT. The probability for a fixed pixel to be randomly assigned as a free pixel is set to be $p = .01$.

For DIPS-LS, again 200 iterations are used, where the first 100 iterations are performed with the soft segmentation and region approach, followed by 100 iterations of boundary refinements (original DART). Again, the initial solution is taken from the SIRT reconstruction, and to solve the intermediate updates (10) and (16), 20 iterations of SIRT are used. For the binary phantoms, the segmentation tolerances were initially set to $r_1 = r_2 = .05$. For phantoms 2 and 3, the tolerances were initially set to $r_1 = r_2 = r_3 = .02$, a smaller choice since the gray values are closer. For all of the phantoms, the criteria for increasing the tolerances was given by (18), where $\varepsilon = .005$, and the increase in the tolerance in each case is set to $\delta = .005$.

For DIPS, 15 iterations of region refinements are used with the update given by (17), followed by 100 iterations of DART. The initial solution for DIPS is computed using TV minimization with the TVAL3 algorithm [26], and the TVAL3 algorithm is also modified and used to also solve each region update step (17). All of the default values are used, except the non-negativity constraint is enforced, and λ_2 is set as $10 \cdot \lambda_1$. The DART iterations are performed in the same way, using 20 iterations of SIRT. The segmentation tolerance r_ℓ and the segmentation increase δ were set consistent with those in the DIPS-LS simulations. Since many fewer iterations are used we choose to relax the condition on increasing the tolerance values that the increase δ is applied whenever (18) is satisfied for $\varepsilon = .1$. As we will see, although we significantly reduce the number of iterations and relax this tolerance due to computational complexity, the model in (17) is powerful enough to still provide promising results.

For DART, DIPS-LS, and DIPS, the smoothing step is computed using a 5×5 Gaussian smoothing kernel, with $\sigma = 2$. We perform several simulations with each reconstruction method for each phantom. First, we test each method whenever the number of available projections is limited, but the angular range of the projections are equally distributed across a full angular projection range of $180°$. With the same projection scheme, we test the methods in the presence of various levels of noise. Last, we test each method whenever the tilt range is limited, a common problem in ET known as the "missing-wedge."

6.1   Limited Data

For this set of experiments, we first use each method when the number of projections available for reconstruction is limited. The projections are taken at equally distributed angles from a full angular range of $180°$. Therefore, the number of projections in which we are limited determines the angular distribution of the projections.

In Figure-4, the rNMP for each approach is shown as a function of the number of projections for the phantoms. The results clearly indicate that each of the discrete tomographic reconstruction methods yield better results than SIRT, except in a few cases whenever the number of projections is most limited and one would likely not be able to trust any reconstruction method. For phantoms 2 and 3, the plot shows that DIPS-LS and DIPS give very similar results, with DIPS showing a slight improvement to DIPS-LS, while both methods yield better results than DART. For phantom 5, the
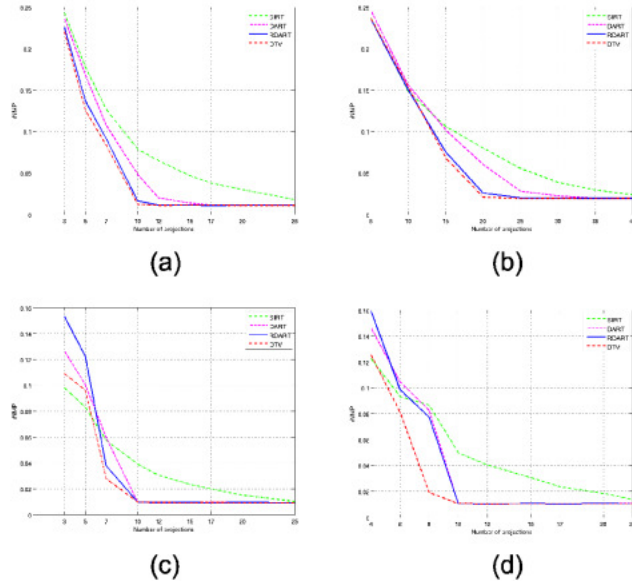
Figure 4: The relative number of misclassified pixels (rNMP) are plotted as a function of the number of available projections.
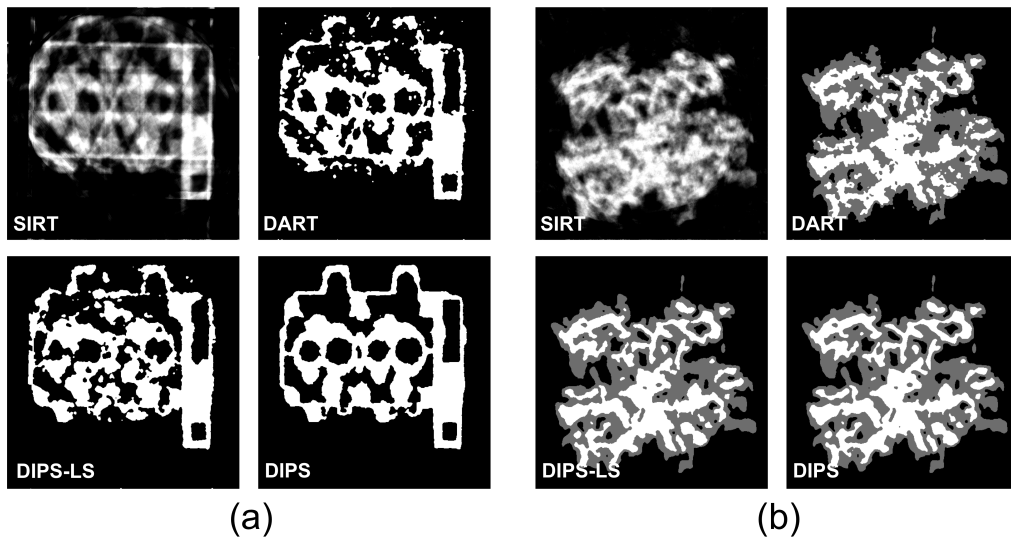

Figure 5: Reconstructions of phantom 5 from 8 projections in (a) and phantom 3 from 20 projection in (a) using a full angular range.
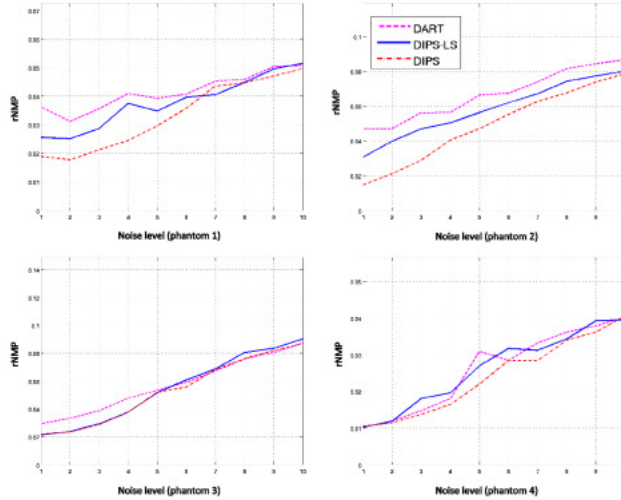
Figure 6: The rNMP is plotted as a function of the noise level, σ.

plot indicates that DART, DIPS-LS, and DIPS all converge to an optimal solution from only 10 projections, and overall DIPS-LS and DART show similar numerical results. However, with fewer projections, DIPS is performing significantly better than all of the methods, and nearly converges to an optimal solution with only 8 projections. Images for these reconstructions are shown in Figure-5, with phantom 3 reconstructed from 20 projections and phantom 5 from 8 projections.

We now add gaussian noise to projection data, where the projections are again taken at equally distributed angles from a full angular range. For each phantom, we fix the number of projections we are using and add various noise levels to the projections. We use normally distributed noise with mean zero and use 10 different values for the standard deviation, $\sigma = 1, 2, \ldots, 10$. In Figure-6, the rNMP is plotted as a function of σ for the phantoms. For a better scaling of the plots only the results for the discrete techniques are shown, since the error from the SIRT reconstruction was significantly greater for each reconstruction.

The results remain more or less consistent with the results from noiseless data, with DIPS generally performing the best, followed by DIPS-LS. The plots for phantoms 1 and 2 show this distinction most clearly, and on the other hand, phantom 4 shows more mixed performance between DIPS-LS and DART. Finally, it should be noted that as the noise level increases, the rNMP for each method appears to be converging to similarly poor results. However, due to the rigid nature of these methods, it is our belief that these techniques are likely not practical with very noisy data.

6.2 Limited Angle

In this set of experiments, the data is again limited, but only by the angular range in which the projections are taken. While we may take as many projections as we desire from the given angular range to create a very large set of data, this type of restriction decreases the variability in the projections, and little or no information is gained by creating additional projections from the same angular range on an already large set of projections.

For each simulation, the projections are taken at every $1°$ from the limited angular range. The angle is restricted to various ranges under a total of $100°$. The angular range is defined to be the difference between the largest angle and the smallest angle.

The plots in Figure-7 show the rNMP as a function of the angular range for phantoms 1, 2, 3, and 5. As with the case of limited projections, SIRT again underperforms compared with
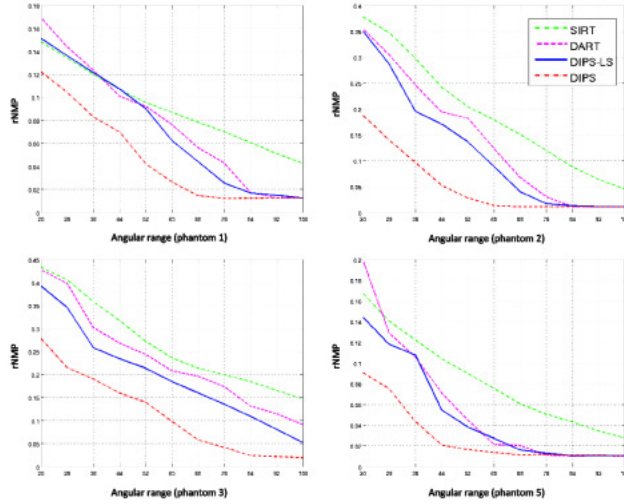
Figure 7: The relative number of misclassified pixels (rNMP) as a function of the angular range.

each of the discrete technique. However, DIPS-LS and DART show mixed results, where DIPS-LS performs consistently better for phantoms 2 and 3, but the other phantoms show only slight improvement with DIPS-LS. Finally, the plots here clearly indicate that DIPS yields significant improvement compared with all other methods. For phantoms 1 and 2, the curve representing the rNMP for DIPS lies consistently below all of the other curves until an optimal solution is reached at an angular range near $60°$, and DART and DIPS-LS reach a near optimal solution nearer to $80°$. Similar arguments can be made for phantoms 3 and 5, where the rNMP curve for DIPS lies constantly below the other curves until an optimal solution is reached.

Figure-8 shows reconstruction images from all approaches for phantoms 2, 3, and 5, from an angular range of $52°$, $76°$, and $44°$ respectively. This limited angle problem is clearly visible in the SIRT reconstructions, showing significant vertical blurring artifacts. The images from the discrete techniques are consistent with the plots, showing slightly improved results with DIPS-LS and significant improvement with DIPS.

In Figure-9, for a demonstration of our method shown are images from various iterations of DIPS for phantoms 1 and 4, with an angular range of $76°$ and $36°$ respectively. The images in Figure-9c highlight the set of free pixels for phantom 4 at each these iterations. These images demonstrate the gradual progression of the solution as it is slowly segmented and refined over the unclassified regions. With each iteration the solution converges closer and closer to a fully discrete solution with fewer free pixels, eventually arriving at a very accurate result. We note that even at the last iteration of DIPS for phantom 4, the set of free pixels has not quite converged to a "boundary-like" set. However, the angular range of only $36°$ is severely limited and such uncertainty should be expected, and whenever the data is less restrictive the set of free pixels will diminish much more rapidly. Even with such a limited range, this reconstruction of phantom 4 still converged to an optimal solution, accurately classifying more than 99% of the pixels with the rNMP = .004898.
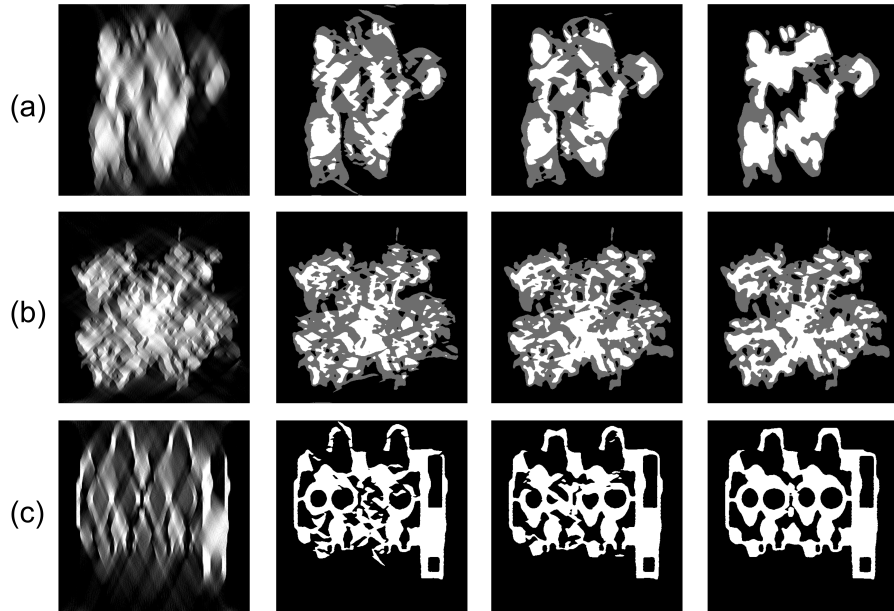
Figure 8: Reconstruction images with all methods. Comparison of reconstruction methods with SIRT (first column), DART (second column), DIPS-LS (third column), and DIPS (fourth column). (a) Phantom 2 using an angular range of 52°. (b) Phantom 3 using an angular range of 76°. (c) Phantom 5 using an angular range of 44°.
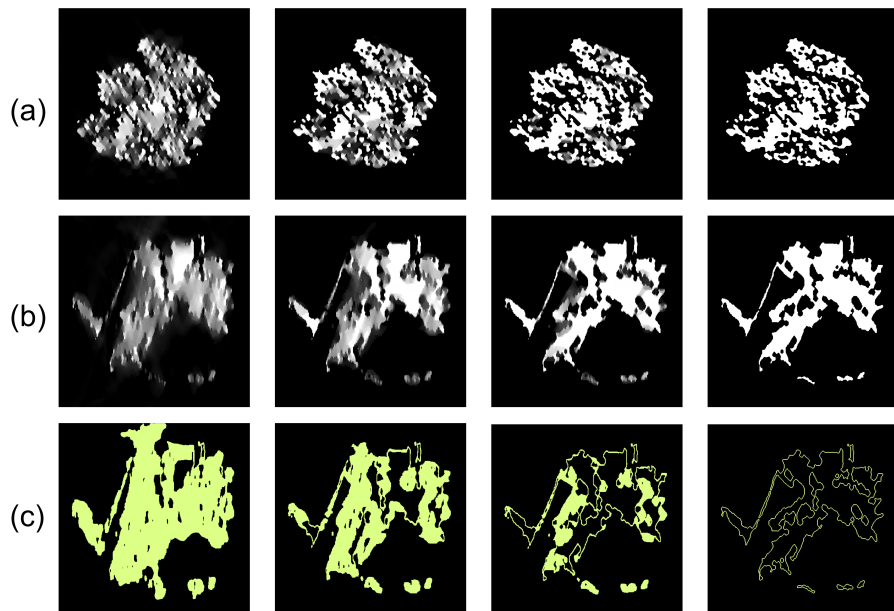

Figure 9: Iterations of DIPS reconstruction method. The iterations shown are the first iteration (first column), the 7th iteration (second column), the 15th iteration just before a full segmentation (third column), and the final result (fourth column). (a) Phantom 1 using an angular range 76°. (b) Phantom 2 using an angular range of 36°. In (c), the set of free pixels for phantom 2 are highlighted at each of these iterations.

7 Conclusion

We have proposed a new discrete reconstruction technique for tomography, which is able to accurately reconstruct objects with only a few compositions. In each iteration, our method combines a new soft partial segmentation technique for careful classification, along with a minimization model that recognizes appropriate sparsities that can be captured through TV regularization. Our simulations already showed improved results with only the partial segmentation approach, which we call DIPS-LS. Furthermore, the more advanced method of DIPS that makes use of TV regularization proved to be the most significant result, yielding far better approximations than the alternative methods. Our simulation results additionally indicate that DIPS offers a significant benefit for electron tomography whenever the angular ranges of the projections are severely limited.

Acknowledgement

References

[1] P. Midgley and M. Weyland. 3d electron microscopy in the physical sciences: the development of z-contrast and eftem tomography. *Ultramicroscopy*, 96(3-4):413–431, SEP, 2003.

[2] S. Marco, T. Boudier, C. Messaoudi, and J. L. Rigaud. Electron tomography of biological samples. *Biochemistry (00062979)*, 69(11):1219–1226, 11 2004.

[3] D. Ropers, U. Baum, K. Pohle, K. Anders, S. Ulzheimer, B. Ohnesorge, C. Schlundt, W. Bautz, W. Daniel, and S. Achenbach. Detection of coronary artery stenoses with thin-slice multi-detector row spiral computed tomography and multiplanar reconstruction. *CIRCULATION*, 107(5):664–666, FEB 11 2003.

[4] F. Natterer. *The Mathematics of Computerized Tomography.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2001.

[5] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. *Journal of Theoretical Biology*, 29:471,IN1,477–476,IN2,481, 1 1970.

[6] P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, 36:105–117, 1 1972.

[7] R. J. Gaudette, D. H. Brooks, C. A. DiMarzio, M. E. Kilmer, E. L. Miller, T. Gaudette, and D. A. Boas. A comparison study of linear reconstruction techniques for diffuse optical tomographic imaging of absorption coefficient. *Physics in Medicine and Biology*, 45(4):1051, 2000.

[8] Wang, Y.L., Yang, J.F., Yin, W.T., & Zhang, Y. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sci.* **1**, 248-272 (2008).

[9] Z. Saghi, *et al.* Three-dimensional morphology of iron oxide nanoparticles with reactive concave surfaces. A compressed sensing-electron tomography approach. *Nano Lett* **11**, 4666-4673 (2011).

[10] B. Goris, W. Van den Broek, K. J. Batenburg, H. H. Mezerji, and S. Bals. Electron tomography based on a total variation minimization reconstruction technique. *Ultramicroscopy*, page 120, 2012.

[11] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010.

[12] G. T. Herman and A. Kuba, Eds., *Discrete Tomography: Foundations, Algorithms, and Applications.* Boston, MA: Birkhaäuser, 1999.

[13] G. T. Herman and A. Kuba, Eds., *Advances in Discrete Tomography and Its Applications.* Boston, MA: Birkhaäuser, 2007.

[14] K. Batenburg. A network flow algorithm for reconstructing binary images from continuous x-rays. *Journal of Mathematical Imaging and Vision*, 30(3):231–248, 2008.

[15] P. Gritzmann, S. deVries, M. Wiegelmann, Approximating binary images from discrete X-rays, *SIAM J. Optim.* 11 (2000) 522Ð546.

[16] A. Frosini, M. Nivat, Binary matrices under the microscope: a tomographical problem, *Theoret. Comput. Sci.* 370 (2007) 201Ð217.

[17] K. J. Batenburg and J. Sijbers. Dart: A practical reconstruction algorithm for discrete tomography. *IEEE Transactions on Image Processing*, 20(9):2542–2553, 09 2011.

[18] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, E. A. Coronado, and G. V. Tendeloo. 3D imaging of nanomaterials by discrete tomography. *Ultramicroscopy*, 109:730–740, 2009.

[19] S. Bals, K. Batenburg, D. Liang, O. Lebedev, G. Van Tendeloo, A. Aerts, J. Martens, and C. Kirschhock. Quantitative three-dimensional modeling of zeotile through discrete electron tomography. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, 131(13):4769–4773, APR 8, 2009.

[20] S. Bals, K. Batenburg, J. Verbeeck, J. Sijbers, and G. Van Tendeloo. Quantitative three-dimensional reconstruction of catalyst particles for bamboo-like carbon nanotubes. *Nano Lett.*, 7(12):3669–3674, DEC, 2007.

[21] Maestre-Deusto, F.J.; Scavello, G.; Pizarro, J.; Galindo, P.L., "ADART: An Adaptive Algebraic Reconstruction Algorithm for Discrete Tomography," *Image Processing, IEEE Transactions on*, **20**, no.8, pp.2146-2152 (2011).

[22] F. Bleichrodt, F. Tabak, K.J. Batenburg, SDART: An algorithm for discrete tomography from noisy projections, *Computer Vision and Image Understanding*, **129**, 63-74 (2014).

[23] Candés, E., Romberg, J., Tao, T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE T. Infor. Theory* **52**, 489-509 (2006).

[24] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS*, 59(8):1207–1223, AUG, 2006.

[25] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, (3):906, 1989.

[26] Li, C. *An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing*, M.A. Thesis, Rice University, US, 2009.