Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

The reconstructed residual error: A novel segmentation evaluation measure for reconstructed images in tomography

Tom Roelandts [a,*], K. Joost Batenburg [a,b,c], Arnold J. den Dekker [a,d], Jan Sijbers [a]

[a] iMinds-Vision Lab, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium
[b] Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands
[c] Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands
[d] Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands

A R T I C L E   I N F O

A B S T R A C T

In this paper, we present the reconstructed residual error, which evaluates the quality of a given segmentation of a reconstructed image in tomography. This novel evaluation method, which is independent of the methods that were used to reconstruct and segment the image, is applicable to segmentations that are based on the density of the scanned object. It provides a spatial map of the errors in the segmented image, based on the projection data. The reconstructed residual error is a reconstruction of the difference between the recorded data and the forward projection of that segmented image. The properties and applications of the algorithm are verified experimentally through simulations and experimental micro-CT data. The experiments show that the reconstructed residual error is close to the true error, that it can improve gray level estimates, and that it can help discriminating between different segmentations.

1. Introduction

In many applications of tomography [1], such as the delineation of anatomical structures (in medical imaging) and object detection (in computer vision), the reconstructed image must be segmented before the results can be analyzed. Segmentation is defined as the classification of image pixels into distinct classes, based on similarity with respect to some characteristic. Numerous methods have been proposed, such as global and local thresholding, region growing, clustering, and atlas-guided approaches [2–4].

Given a segmentation, objectively evaluating the accuracy of that segmentation is not a trivial task [5,6]. *Supervised* methods evaluate a segmentation algorithm by comparing its output with *gold standard* segmentations. However, since such segmentations are typically not available in practice, they must often be generated manually, which is not easy and may make the evaluation subjective. *Unsupervised* methods (also sometimes called *stand-alone methods* [7]) do not need gold standards, as they evaluate the segmentation results directly, using one or more of the *de facto* standard criteria of Haralick and Shapiro [2]. These methods are objective and applicable to a wide variety of images, but their analysis is restricted to the segmentation result itself. A possibility that is often overlooked, in the specific case of tomography, is

exploiting the available projection images, which can provide external information about the segmented image.

The current paper introduces the *reconstructed residual error*, which *does* exploit the original projection images. Our method is applicable to reconstruction problems for which the segmentation is based on the density of the scanned object, where we use the term *density* to refer to the particular physical property of the object of which linear projections are acquired during the scanning process (e.g., mass density, X-ray attenuation, electron beam scattering, etc.). The reconstructed residual error is an unsupervised evaluation in the terminology of [6], since it is an objective evaluation at the level of the segmentation itself that does not need a reference image [6, Fig. 1]. In contrast to the unsupervised methods surveyed in [6], the proposed method does not have to rely on the criteria of Haralick and Shapiro, since it uses the projection images as external information.

The reconstructed residual error evaluates a given segmentation by providing a spatial map of the errors. It is computed by reconstructing the difference between the recorded data and the forward projection of that segmentation. The computation of the error map is independent of the methods that were used for reconstructing the image and determining the segmentation.

The remainder of this paper is organized as follows. In Section 2, the reconstructed residual error is defined and its properties are described in detail. Section 3 reports on the results of experiments, using both simulations and experimental micro-CT data. These

---

* Corresponding author.
E-mail address: tom.roelandts@uantwerpen.be (T. Roelandts).

results are discussed in Section 4, and conclusions are drawn in Section 5.

2. Reconstructed residual error

Here, we describe the reconstructed residual error. We first present an intuitive overview of its computation, before giving a complete description of its properties.

2.1. Overview

Fig. 1 presents an overview of the computation of the reconstructed residual error. The example is based on a two-dimensional slice through objects with simple geometrical shapes and only two different gray levels (Fig. 1a). The original projection data (Fig. 1b) is a sinogram, acquired by rotating around the object. This sinogram is then reconstructed (Fig. 1c) and segmented (Fig. 1e). For this overview figure, the segmented reconstruction was computed by thresholding Fig. 1c, and subsequently choosing gray levels. The reconstructed residual error is computed from the original projection data (Fig. 1b) and the segmented reconstruction (Fig. 1e). The original (non-segmented) reconstruction (Fig. 1c) is not used.

To compute the reconstructed residual error, the segmented reconstruction (Fig. 1e) is first projected forward. The result of this operation (Fig. 1d) is then subtracted from the original projections, resulting in the *residual projection error* (Fig. 1g). The residual projection error is then reconstructed to provide the reconstructed residual error (Fig. 1h), which provides a spatial map of the segmentation error. From Fig. 1h, it is clear that both gray levels of the segmented reconstruction are incorrect. The erroneous lines and dots in the segmented reconstruction (Fig. 1e), which are caused by artifacts in the reconstruction (Fig. 1c), are also clearly visible in Fig. 1h. Note that the true error (Fig. 1i), which is the difference between the original object (Fig. 1a) and the segmented reconstruction, is quite close to the reconstructed residual error.

An alternative that might be considered, is to simply compute the difference between the original and the segmented reconstruction (Fig. 1f). However, this difference can be expected to show much more reconstruction artifacts, as is explained in Section 2.4. An intuitive way to see this is that, for phantom experiments, the segmented reconstruction can potentially be identical to the phantom, in which case the residual projection error (Fig. 1g) would be zero. Hence, a perfect segmentation would result in the reconstructed residual error (Fig. 1h) being zero everywhere, while the difference between the original and the segmented reconstruction (Fig. 1f) will always exhibit the reconstruction artifacts of the original reconstruction (Fig. 1c).

2.2. Notation and concepts

The projection process in tomography can be modeled as a linear operator that is determined by the projection geometry. This leads to a system of linear equations,

$$Wx + n = \tilde{p},\qquad(1)$$

where $\tilde{p} \in \mathbb{R}^m$ is the measured projection data, $n \in \mathbb{R}^m$ is the noise, and $x \in \mathbb{R}^n$ is the unknown image. The linear operator is represented by the $m \times n$ matrix W, the *projection matrix*. An approximate solution $\hat{x} \in \mathbb{R}^n$ of (1) can then be computed, for example by minimizing some norm $\|Wx - \tilde{p}\|$.

In Sections 2.4 and 2.5, we assume that $n = 0$ and explicitly refer to the ground truth object as $g \in \mathbb{R}^n$, which leads to $p = Wg$, where $p \in \mathbb{R}^m$ are noiseless projections. We denote a general reconstruction algorithm as an operator $\mathcal{R}: \mathbb{R}^m \to \mathbb{R}^n$, which leads to $\hat{x} = \mathcal{R}Wg$. The *so-called* shortest least squares solution is denoted as x^+, where $x^+ = W^+p$, with W^+ the Moore–Penrose pseudoinverse [8, Section 7.3] of W. We denote the *row* (or *image*) *space* of W by $R(W)$, and the *null space* by $N(W)$. See again [8] for details on these concepts from linear algebra.
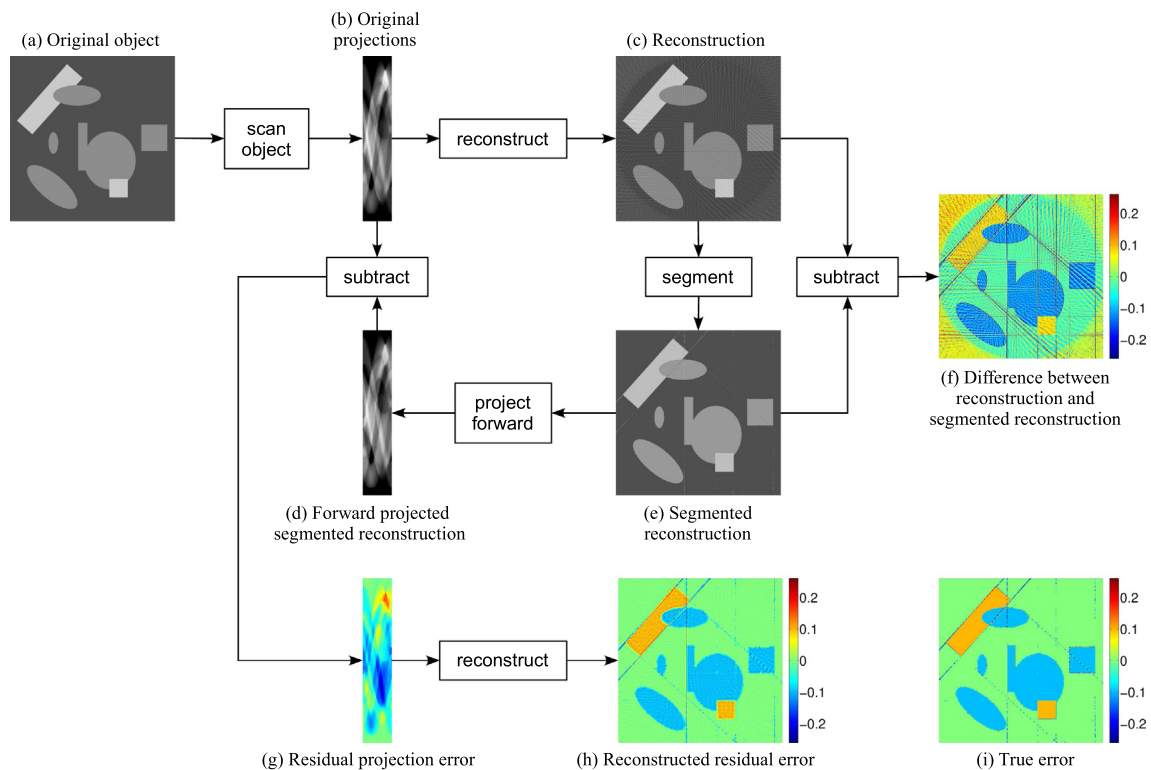

Fig. 1. Overview of the computation of the reconstructed residual error.

A segmentation partitions the pixels of an image into sets $Y_1, \ldots, Y_d$, where d is the number of classes (segments) in the segmented image. Determining d can be a difficult problem [9]. However, since the proposed method operates on an already segmented image, we can assume d to be known. If it is assumed that the segmentation is based on the density of the scanned object, then it is possible to assign a gray level ρ_k to each class Y_k that is also related to that density. In Section 2, we assume that segmentation has been performed and gray levels have been assigned, resulting in a segmented image $s \in \mathbb{R}^n$. The actual segmentation algorithm and the way in which the gray levels were selected is not relevant for the definition and the properties of the proposed method. In Section 3, we *do* detail how we performed these operations in the experiments.

2.3. Definition

Here, we define the reconstructed residual error. Inputs are the segmented image s and the measured projection data $\tilde{p}$. First, s is forward projected to give $p_s \in \mathbb{R}^m$, so $p_s = Ws$. The *residual projection error* $p_e \in \mathbb{R}^m$ is then defined as $p_e = \tilde{p} - p_s$. Second, the residual error is reconstructed by computing an approximate solution of the system

$$We = \tilde{p} - p_s, \tag{2}$$

as for the system of (1). Here, $e \in \mathbb{R}^n$ corresponds to the (unknown) error image. The approximate solution $\hat{e} \in \mathbb{R}^n$ is then defined to be the reconstructed residual error. The least squares solution is again denoted by e^+, as for x^+.

The theoretical treatment of the reconstructed residual error in Sections 2.4 and 2.5 is applicable in two and three dimensions, and is independent of the geometry of the scanner (e.g., parallel beam, cone beam, etc.), since that simply results in a different W in (1) and (2).

2.4. Least squares approach

In this section, we study the reconstructed residual error under the assumption that the Moore–Penrose pseudoinverse W^+ is used, both for the original reconstruction and to compute the reconstructed residual error itself. Doing so makes the reconstruction algorithm a well-defined operator, and allows us to be mathematically rigorous. However, despite being useful as a mathematical model, the Moore–Penrose pseudoinverse is not a practical reconstruction method due to, e.g., its susceptibility to noise and high computational cost. Therefore, inverting the projection matrix is not attempted in practice, and iterative methods are used to solve the reconstruction problem [10]. Hence, after our initial analysis using the pseudoinverse, we switch to practical reconstruction methods in Section 2.5 and show that these also result in an error map that is close to the true one.

First, we formally define what we mean by the term *artifact*. Typically, a reconstructed image contains different types of artifacts, with causes such as scanner imperfections, noise, the reconstruction algorithm, and the segmentation algorithm. In the current paper, we consider all differences between the ground truth object and the (possibly segmented) final image to be artifacts.

Definition 1. The artifacts of an image $y \in \mathbb{R}^n$ are $g - y$, where $g \in \mathbb{R}^n$ is the ground truth object.

Each image $y \in \mathbb{R}^n$ can be split into two orthogonal components, with the first in the row space and the second in the null space of W. We write this as $y = y_R + y_N$, with $y_R, y_N \in \mathbb{R}^n$, and $y_R \perp y_N$.

Below, we reiterate a few basic properties of W and W^+, for convenience.

Property 1: $R(W) \perp N(W)$.
Property 2: $WW^+W = W$.
Property 3: $W^+Wy = y_R$.

Property 3 implies that the shortest least squares reconstruction x^+ exactly recovers the row space component of g, since $x^+ = W^+p = W^+Wg = g_R$.

Proposition 1. $g - x^+ \in N(W)$, *i.e., the artifacts of the shortest least squares solution are in the null space of W.*

We have $x^+ = W^+p = W^+Wg$. It is instructive to split x^+ into the ground truth and a component that represents the artifacts, as $x^+ = W^+Wg = g + (W^+W - I)g$. On forward projection, the second component disappears, since $WW^+W - W = 0$. Hence, $Wx^+ = Wg$, and the artifacts are indeed in $N(W)$.

The implication of Proposition 1 is that the reconstruction artifacts of x^+ cannot be studied or further reduced by comparing them with the projection data p after forward projection.

This changes when the reconstruction is segmented. Reconstruction artifacts are typically non-homogeneous structures such as streaks and blurring. Segmentation into regions that each have a given gray level ρ_k, however, results in an image that is composed of homogeneous regions, since the segmentation removes the small variations of the gray levels within each region. Hence, it removes a large part of the reconstruction artifacts, while, of course, also introducing new artifacts. These new artifacts are typically very different from reconstruction artifacts, and mainly consist of erroneous gray levels and wrongly classified pixels at the edges of different regions in the image. In contrast with reconstruction artifacts, which are in the null space of W, segmentation artifacts are not expected to have zero ray sums. In a homogeneous region with a gray level offset, individual pixels all share the same error, resulting in a nonzero total contribution. This means that the artifacts of s are expected to have a (large) component in $R(W)$, and *can* be studied and further reduced by comparing them with the projection data after forward projection. The reconstructed residual error is an application of this.

Denote the segmentation artifacts $a \in \mathbb{R}^n$ of s as $a \equiv g - s$. We then have the following proposition.

Proposition 2. $e^+ = a_R$, *i.e., the reconstructed residual error exactly recovers the row space component of a.*

From (2) (with noiseless p substituted for $\tilde{p}$) and $p_s = Ws$, we have $e^+ = W^+(p - p_s) = W^+Wg - W^+Ws = W^+Wa = a_R$.

This means that the reconstructed residual error is a reconstruction of the segmentation artifacts in the same way that the original image is a reconstruction of the scanned object, i.e., we have $e^+ = a_R$ and $x^+ = g_R$. The main implication of Proposition 2 is that the reconstructed residual error can be used to study the segmentation artifacts directly, starting from the segmented reconstruction.

Finally, we briefly return to the alternative of using the difference between the reconstruction and the segmented reconstruction, $x^+ - s$, as an estimate of the segmentation artifacts a, as mentioned in Section 2.1. We compare the artifacts of $x^+ - s$ with those of e^+. The artifacts of e^+ are $a - e^+$. The artifacts of $x^+ - s$, seen as an estimate of a, are exactly the artifacts of x^+, since $a - (x^+ - s) = g - s - x^+ + s = g - x^+$. And, since $\|a\| \ll \|g\|$, we typically expect that $\|a - e^+\| \ll \|g - x^+\|$, which would imply that the artifacts of $x^+ - s$ make it a poor estimate of a.

2.5. General reconstruction algorithms

Leaving behind the "perfect" reconstruction algorithm that is the Moore–Penrose pseudoinverse, we now investigate how the results from Section 2.4 change when W^+ is replaced by a practical reconstruction algorithm. In Section 3, we show experimentally that doing so still results in useful maps of the segmentation artifacts. There, we employ several different reconstruction algorithms, not restricted to linear ones for the reconstruction of the original image.

If a general, possibly non-linear, reconstruction algorithm is used for the initial reconstruction, then Proposition 1 is no longer applicable. However, the reconstructed residual error is computed from a *segmented* reconstruction s, which is still expected to have a (large) component in $R(W)$. Hence, the effectiveness of the reconstructed residual error is not affected. Moreover, depending on the component of the artifacts that is in $R(W)$, it might be possible to study the artifacts of the original reconstructed image directly, without a separate segmentation step. An obvious example of this are discrete algorithms such as DART (*Discrete Algebraic Reconstruction Technique*), which directly produce a fully segmented result [11].

A general reconstruction algorithm can also replace W^+ for the computation of the reconstructed residual error itself. The result is then no longer exactly the row space component of a. We have the following proposition.

Proposition 3. *For a general reconstruction algorithm $\mathcal{R}$, $\hat{e} = \mathcal{R}Wa$, i.e., the reconstructed residual error is the reconstruction of the segmentation artifacts a.*

From (2) (with noiseless p substituted for $\tilde{p}$) and $p_s = Ws$, we have $\hat{e} = \mathcal{R}(p - p_s) = \mathcal{R}(Wg - Ws) = \mathcal{R}Wa$.

Proposition 3 states that $\hat{e}$ is a reconstruction of a in the same sense that $\hat{x}$ is a reconstruction of the ground truth, since $\hat{x} = \mathcal{R}Wg$. Together with the shortest least squares solution often not being a desirable solution anyway, due to noise in the projection data, this implies that W^+ can be replaced by a suitable more general reconstruction algorithm for computing the reconstructed residual error.

3. Experiments and results

In this section, we describe the experiments, for both simulated and experimental data, that were carried out to investigate the properties of the reconstructed residual error. To improve the clarity of the presented results, the simulation experiments are based on two-dimensional slices and parallel-beam geometry. However, as was already mentioned in Section 2.3, the method is readily applicable to three-dimensional objects and other geometries such as cone beam. The application in Section 3.3 uses cone-beam geometry. Since computing the reconstructed residual error is essentially a reconstruction problem, the quality of the result depends on the number of available projections. To demonstrate that the algorithm is also useful when only relatively few projections are available, most experiments were based on synthetic datasets with only 90 projections.

Three *phantom images* were created, the first one (Fig. 2a) is of size 64×64 pixels, to allow exact least squares computations using W^+. The experiment using Phantom 1 was performed on a grid of size 64×64 pixels. Phantoms 2 and 3 (Fig. 2b and c) are of size 2048×2048 pixels. These phantoms have a higher resolution than the reconstruction grid, which is of size 512×512 pixels, to reduce the effect of the pixelation on the reconstructions.

(a) Phantom 1

(b) Phantom 2   (c) Phantom 3
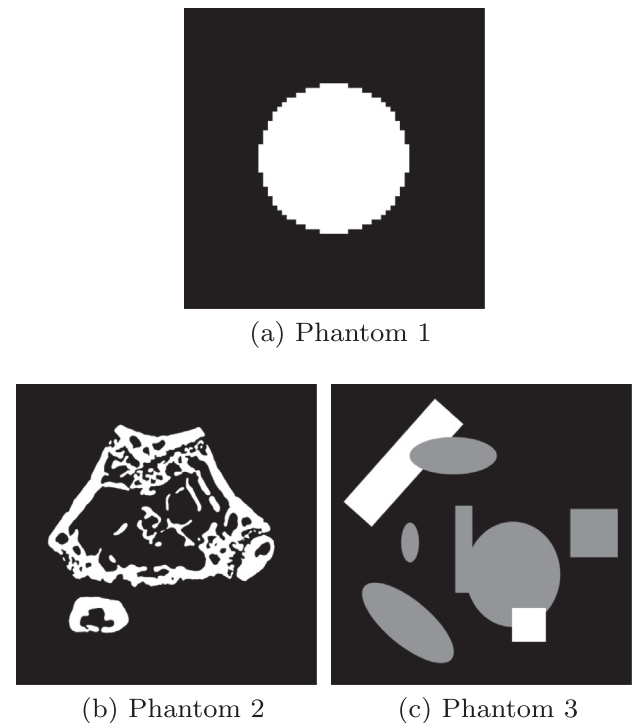
Fig. 2. Phantom images. (a) Two gray levels, 64×64 pixels. (b) Two gray levels, based on the FBP reconstruction of a mouse femur. (c) Three gray levels.

Below, we provide a brief overview of the reconstruction algorithms that were used in the experiments.

(1) FBP (*Filtered Backprojection*) is an analytical technique that is directly related to the inverse Radon transform [1]. A standard implementation of FBP was used, with a Ram–Lak filter and linear interpolation in the projection domain.

(2) CGLS (*Conjugate Gradient Least Squares*) is an algebraic technique that computes the shortest least squares solution of (1) [12]. We used 300 iterations of CGLS to approximate the effect of W^+, allowing for the use of a larger reconstruction area than is possible when using the pseudoinverse directly.

(3) SIRT (*Simultaneous Iterative Reconstruction Technique*) is an algebraic technique that computes a *weighted* least squares solution of (1) [13]. SIRT was implemented as defined in [14], performing 300 iterations.

(4) PDART [15] (*Partially Discrete Algebraic Reconstruction Technique*) is a partially discrete technique that assumes that only the densest material is homogeneous. The gray level of the densest material is exploited as prior knowledge. A total of 300 iterations was performed.

We used the well-known global thresholding method by Otsu [16] to compute all segmentations. Otsu's method provides a threshold (or several thresholds for the multi-level version of the algorithm) that can then be used to assign each pixel of the image to a segmentation class Y_k. If appropriate gray levels $\rho_1, \ldots, \rho_d$ are known, e.g., through prior knowledge about the scanned object, they can then be assigned to the pixels in the sets $Y_1, \ldots, Y_d$, resulting in the segmented image $s \in \mathbb{R}^n$. We did not assume such prior knowledge in the experiments, and have used the mean of all pixel values of the original reconstruction that share a segmentation class as an initial estimate,

$$\hat{\rho}_k = \frac{1}{|Y_k|} \sum_{y \in Y_k} y, \quad \text{for each } k \in \{1, \ldots, d\}. \tag{3}$$

In the experiments of Sections 3.2 and 3.3, the reconstructed residual error itself is used to improve upon these initial estimates.

For all experiments, the reconstructed residual error was computed using 300 iterations of SIRT, to make the results comparable.

To determine numerically how close a given error image $e \in \mathbb{R}^n$ is to the true error, we used the *relative Euclidean distance* $d : \mathbb{R}^n \to \mathbb{R}$, defined as

$$d(e) = \frac{\|e - e_t\|_2}{\|e_t\|_2}, \tag{4}$$

where $e_t \in \mathbb{R}^n$ is the true error.

We now provide an overview of the experiments that follow. First, the basic properties of the algorithm are studied (Section 3.1). Then, two applications are presented. It is demonstrated that the technique can be used to detect and correct errors in the estimate of the gray levels (Section 3.2). Finally, the technique is applied to experimental micro-CT data (Section 3.3), to interpret the differences between two segmentations.

3.1. Properties of the algorithm

The first experiment demonstrates the least squares approach from Section 2.4. Both the original and the segmented reconstruction were split into their row and null space components. To be able to compute W^+ in practice, a small reconstruction area of size 64×64 pixels was used, and only 32 projections were taken. This results in a system matrix W of size 2048×4096, and W^+ of size 4096×2048. A synthetic dataset was created from Phantom 1 (Fig. 2a), also shown in Fig. 3d, using 32 equiangular parallel beam projections. Different from the other simulation experiments, noise was not added to this dataset. The reconstruction (Fig. 3e) was then computed using the Moore–Penrose pseudoinverse W^+, making it the exact shortest least squares solution. The reconstruction was segmented using Otsu's method (Fig. 3g). The gray levels of the segmented reconstruction were determined using (3) on the pixels

of the original reconstruction, with the Otsu segmentation classes as Y_k ($k = 1, 2$).

The artifacts of the reconstruction are shown separately (Fig. 3f), and split into row and null space components. As shown theoretically by Proposition 1, the artifacts are completely in the null space component (Fig. 3j). The row space component (Fig. 3a) is zero everywhere. The artifacts of the segmented reconstruction are also shown separately (Fig. 3h). The main assumption that was made in Section 2.4, was that the segmentation artifacts have a (large) component in the row space of W. This assumption is confirmed by again splitting these artifacts into row and null space components. The null space component (Fig. 3k) is not zero, but most of the artifacts are clearly in the row space component (Fig. 3b). Using the least squares approach of Section 2.4, this row space component is exactly e^+. However, when the reconstructed residual error is computed using SIRT (Fig. 3c), the result is very close to e^+, indicating that SIRT is an acceptable practical alternative for W^+ in this case. The true error (Fig. 3m), which is the difference between the original object (Fig. 3d) and the segmented reconstruction (Fig. 3g), is close to both e^+ and the reconstructed residual error as computed with SIRT. The difference between the original and the segmented reconstruction (Fig. 3i) can clearly not be used as an alternative. This is confirmed numerically through the first row of Table 1, where (A) is the distance, as defined by (4), between Fig. 3b and m, (B) is the distance between Fig. 3c and m, and (C) is the distance between Fig. 3i and m.

The second experiment makes the transition from the least squares approach to a more practical approach on a larger reconstruction grid. For this experiment, a synthetic dataset was created from Phantom 2 (Fig. 2b), using 90 parallel beam projections, evenly spaced at $2°$ intervals. A detector with 512 pixels was used, to simulate the practical situation where the detector pixel size equals the width of the reconstruction grid. Poisson noise, corresponding to 10^5 initial photons per detector pixel, was applied to the synthetic dataset. The dataset was then reconstructed using 300 iterations of SIRT (Fig. 4a), and segmented using Otsu's method (Fig. 4d). The gray levels of Fig. 4d were determined using (3) on the pixels of Fig. 4a, with the Otsu segmentation classes as Y_k ($k = 1, 2$).
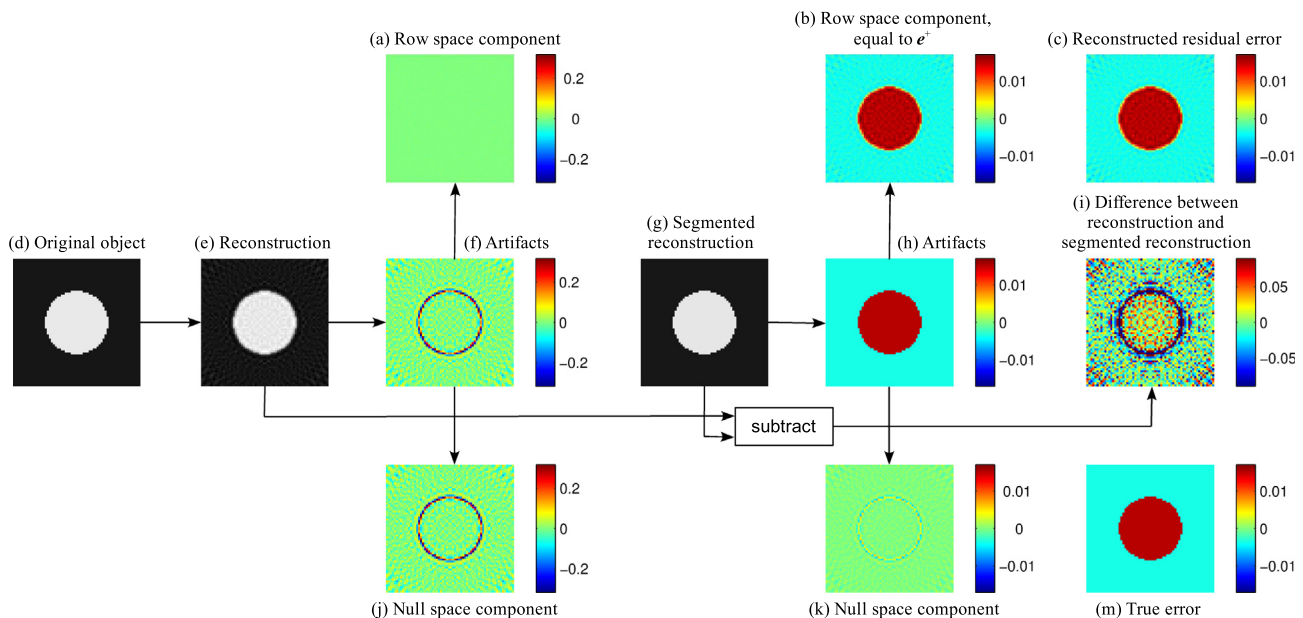

Fig. 3. Complete results for a simple phantom, using a least squares approach. Artifacts of both the original and the segmented reconstruction are split into row and null space components.

Table 1
Distance (4) between true error and (A) row space component of artifacts, (B) reconstructed residual error computed with SIRT, and (C) difference between original and segmented reconstructions. The first column shows the phantom and algorithm that were used for the original reconstruction.
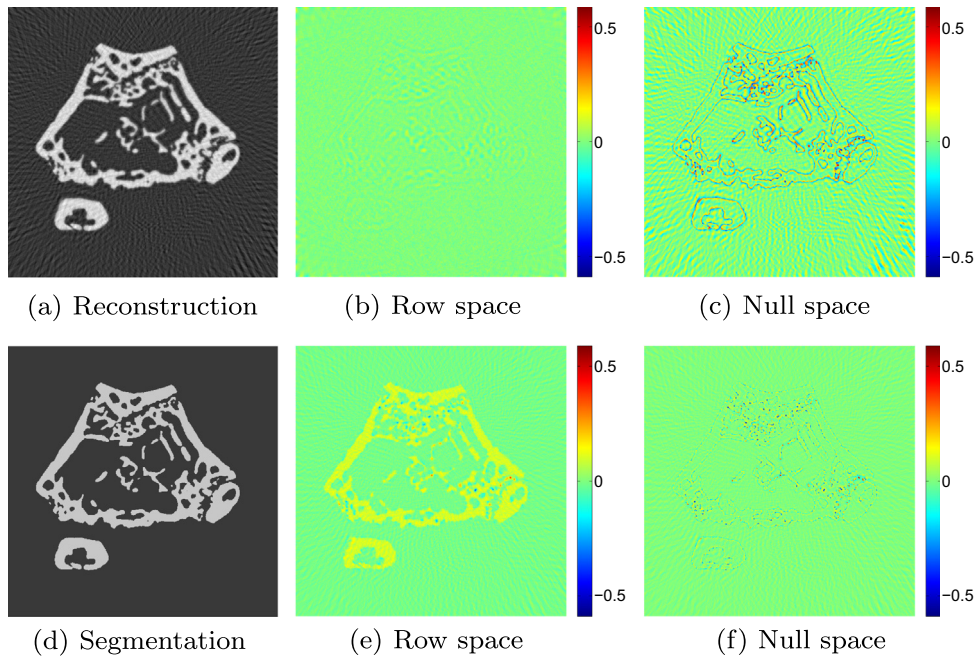
Reconstruction	A	B	C
Phantom 1, W^+	0.14	0.15	7.31
Phantom 2, SIRT	0.67	0.81	1.98
Phantom 3, FBP	0.66	0.95	3.14
Phantom 3, SIRT	0.80	1.31	2.32

Using 300 iterations of CGLS as an approximation of W^+, the artifacts of both the original and the segmented reconstruction were then split into their row and null space components. The artifacts of the reconstruction are again mainly in the null space component (Fig. 4c), even though SIRT has replaced W^+ as the reconstruction algorithm. The row space component (Fig. 4b) is small. For the artifacts of the segmented reconstruction, this is clearly different. The gray level error of the "bone" material, which

is quite subtle because the gray levels were computed from the original reconstruction, clearly shows as an artifact in the row space component (Fig. 4e). The null space component (Fig. 4f) is small.

If the reconstructed residual error is computed using SIRT (Fig. 5a), then the result is close to the row space component of the segmentation artifacts (Fig. 4e). The reconstructed residual error is also close to the true error (Fig. 5b). As before, Table 1 (second row) numerically confirms that both the row space component (A) and the reconstructed residual error (B) are closer to the true error than the difference between the original and the segmented reconstruction (C).

The third experiment demonstrates that relatively subtle differences between segmented reconstructions can be visualized using the reconstructed residual error. This experiment is based on Phantom 3. A synthetic dataset was created, again using 90 parallel beam projections. Poisson noise, again corresponding to 10^5 initial photons per detector pixel, was applied. The dataset was then reconstructed twice, once using FBP (Fig. 6a), and once using 300

(a) Reconstruction (b) Row space (c) Null space

(d) Segmentation (e) Row space (f) Null space

Fig. 4. (a) SIRT reconstruction of Phantom 2. (b) Row space component of the reconstruction artifacts. (c) Null space component of the reconstruction artifacts. (d) Segmented SIRT reconstruction of Phantom 2. (e) Row space component of the segmentation artifacts. (f) Null space component of the segmentation artifacts.

(a) Reconstructed error (b) True error
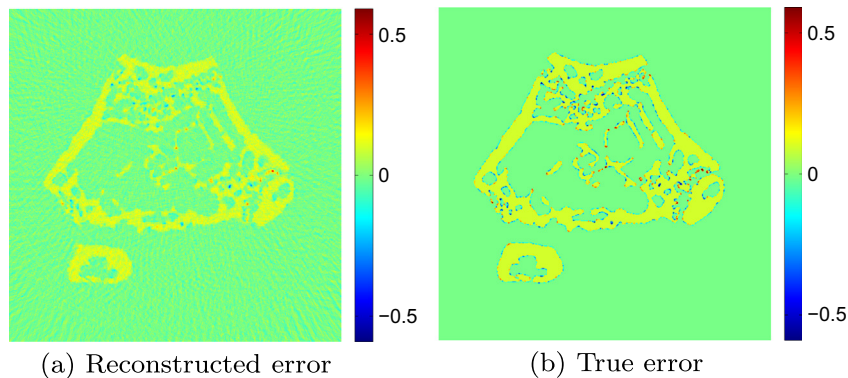
Fig. 5. (a) Reconstructed residual error for the segmented reconstruction of Phantom 2. (b) True error.

iterations of SIRT (Fig. 6d). The gray levels were again computed from the original reconstruction, using the Otsu segmentation classes.

As before, the reconstructed residual error (Fig. 6b and e) is close to the true error (Fig. 6c and f, respectively). This is again confirmed numerically through Table 1 (third and fourth row). The gray level errors are apparently quite small, in particular for the segmented SIRT reconstruction (Fig. 6e). However, as we show numerically in Section 3.2, Fig. 6e can still be used to correct the gray levels from Fig. 6d.

3.2. Application—improving gray level estimates

In this section, we demonstrate that the gray level estimates of a segmented reconstruction can be improved using the reconstructed residual error. Visually, Figs. 3c, 5a, and 6b already suggest that the gray levels of the different objects are not correct, since there is a clear and relatively uniform error inside the objects. In Fig. 6e, this effect is less obvious, but Table 2 shows that this result can still be used to correct the gray levels.

Table 2 provides a numerical overview of the corrections that were achieved for the experiments of Section 3.1. The true gray levels ρ_k were taken from the phantoms (Fig. 2), with $k = 1, 2$ for Phantoms 1 and 2, and $k = 1, 2, 3$ for Phantom 3. The estimated gray levels $\hat{\rho}_k$ are the gray levels of the segmented reconstructions, computed using (3) from the original reconstructions. The estimated gray level errors $\hat{\varepsilon}_k$ were computed from the reconstructed residual error using (3), the same procedure as for the $\hat{\rho}_k$. The sum of $\hat{\rho}_k$ and $\hat{\varepsilon}_k$ is shown as the corrected gray level $\hat{\rho}'_k$.

From comparison of the corrected gray levels $\hat{\rho}'_k$ with the true gray levels ρ_k from the phantom, it is clear that $\hat{\varepsilon}_k$ is a good estimate of the difference between $\hat{\rho}_k$ and ρ_k, and that it can be used to correct $\hat{\rho}_k$. The corrections that are applied to the gray levels are relatively small. However, they still have a visible effect on the segmented reconstructions. This is clear from the reconstructed residual error (Fig. 7) for reconstructions in which the corrected gray levels were used, when Figs. 7a–c are compared with Figs. 5a, 6b, and 6e, respectively.

Table 2
Improving gray level estimates, showing true gray levels (ρ_k), estimated gray levels ($\hat{\rho}_k$), estimated gray level errors ($\hat{\varepsilon}_k$), and corrected gray levels ($\hat{\rho}'_k$). The first column shows the phantom and algorithm that were used for the original reconstruction.

| Reconstruction	Param.	$k = 1$	$k = 2$	$k = 3$
Phantom 1, $\boldsymbol{W}^+$	ρ_k	0.000	1.000	
	$\hat{\rho}_k$	0.004	0.985	
	$\hat{\varepsilon}_k$	−0.004	0.015	
	$\hat{\rho}'_k$	0.000	1.000	
Phantom 2, SIRT	ρ_k	0.000	1.000	
	$\hat{\rho}_k$	0.012	0.904	
	$\hat{\varepsilon}_k$	−0.011	0.071	
	$\hat{\rho}'_k$	0.000	0.975	
Phantom 3, FBP	ρ_k	0.000	0.502	1.000
	$\hat{\rho}_k$	−0.005	0.474	0.973
	$\hat{\varepsilon}_k$	0.004	0.024	0.024
	$\hat{\rho}'_k$	−0.001	0.498	0.998
Phantom 3, SIRT	ρ_k	0.000	0.502	1.000
	$\hat{\rho}_k$	0.003	0.495	0.992
	$\hat{\varepsilon}_k$	−0.003	0.006	0.010
	$\hat{\rho}'_k$	−0.000	0.502	1.002

3.3. Application—discriminating between segmentations

In this experiment, the reconstructed residual error was computed for experimental cone beam data. A dataset of an ex vivo mouse femur was acquired with a SkyScan 1172 micro-CT scanner using 376 projections at $0.5°$ intervals, at a detector resolution of 5 μm. The SkyScan NRecon software package was used to correct for ring and beam-hardening artifacts.

The dataset was reconstructed using two different algorithms. First, it was reconstructed using 300 iterations of SIRT (Fig. 8a) and segmented using Otsu's method (Fig. 8b). The gray levels of the segmented reconstruction were estimated using (3) and the SIRT reconstruction. Second, it was reconstructed using 300 iterations of PDART (Fig. 9c), using the gray level from the bone material of the segmented SIRT reconstruction as prior knowledge.

(a) Segm. FBP (b) Reconst. error (c) True error

(d) Segm. SIRT (e) Reconst. error (f) True error
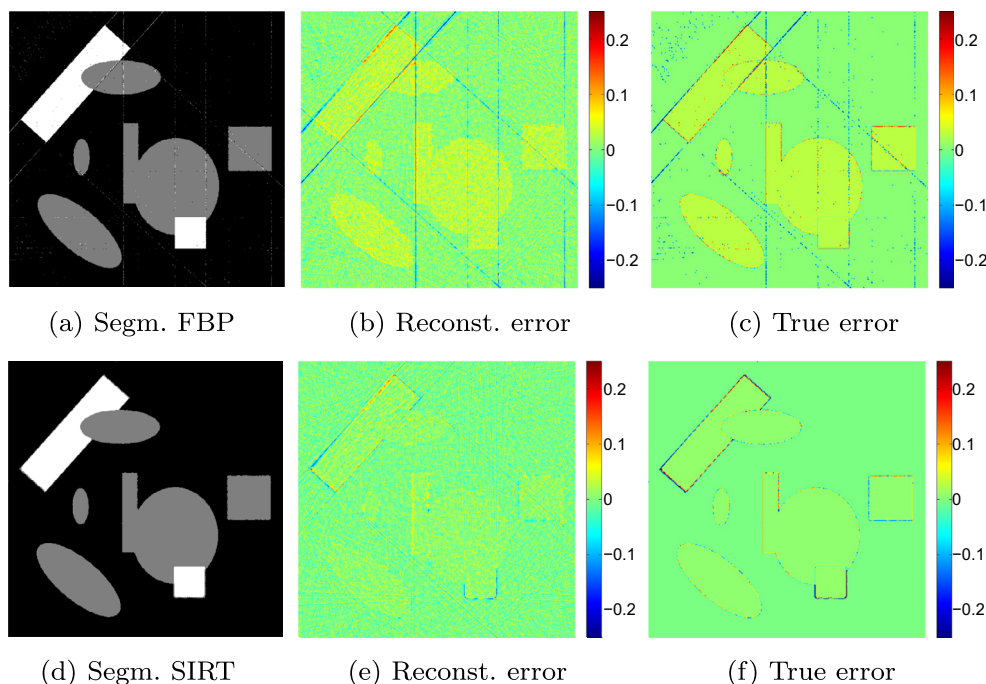
Fig. 6. (a) Segmented FBP reconstruction of Phantom 3. (b) Reconstructed residual error of (a). (c) True error of (a). (d) Segmented SIRT reconstruction of Phantom 3. (e) Reconstructed residual error of (d). (f) True error of (d).

(a) Phantom 2, SIRT



(b) Phantom 3, FBP
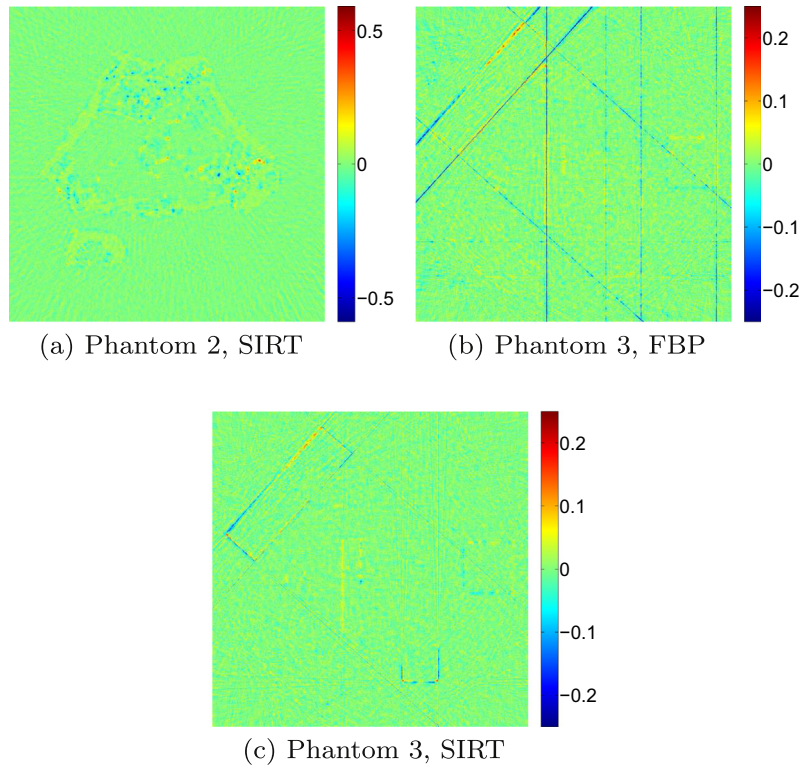
(c) Phantom 3, SIRT

Fig. 7. Reconstructed residual error for reconstructions that use corrected gray levels. (a) Phantom 2, using SIRT. (b) Phantom 3, using FBP. (c) Phantom 3, using SIRT.

(a) SIRT

(b) Segmented SIRT

(c) PDART, step 1

(d) PDART, step 6
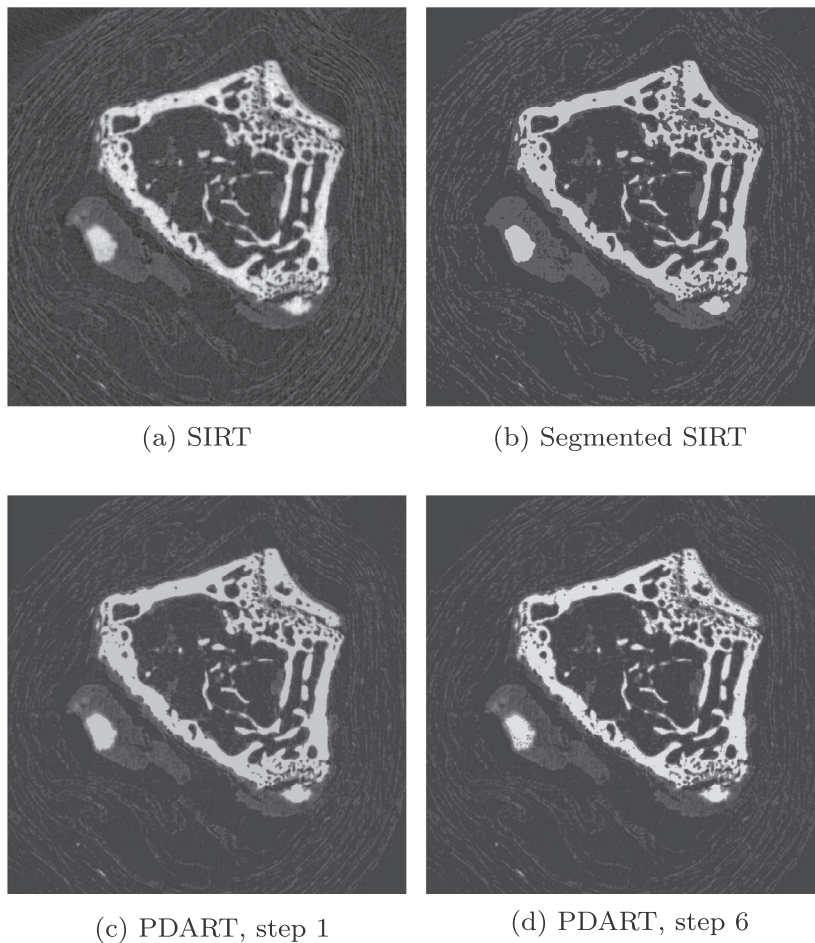
Fig. 8. (a) SIRT reconstruction of a single slice of the femur dataset. (b) The reconstruction from (a), segmented using Otsu's method. (c) PDART reconstruction, using the initial gray level estimate. (d) PDART reconstruction after six correction steps.

(a) SIRT, initial

(b) SIRT, corrected

(c) PDART, step 1

(d) PDART, step 6
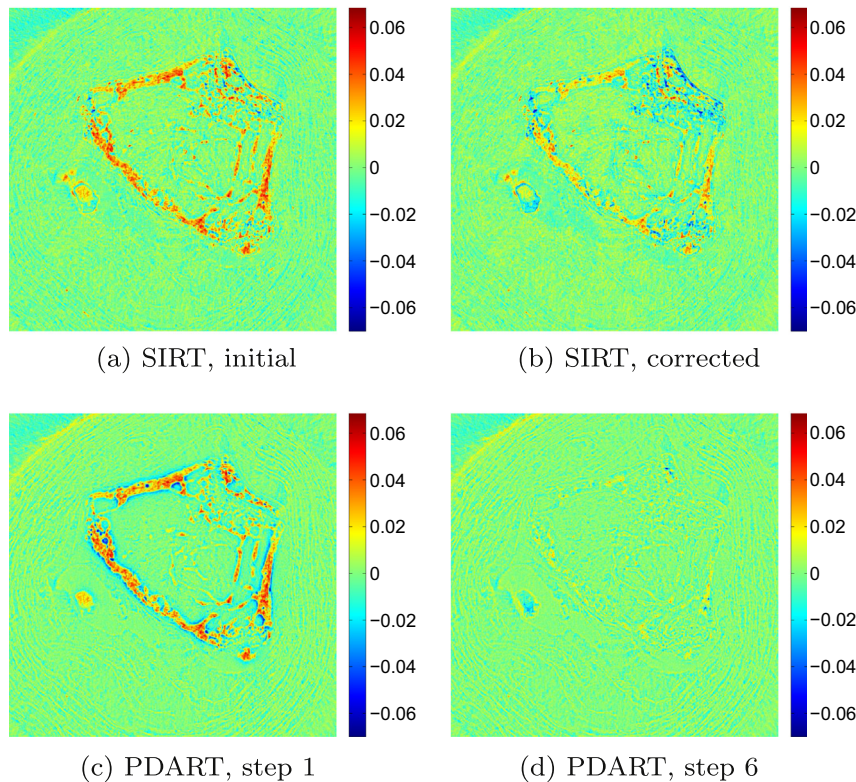
Fig. 9. Reconstructed residual error for the femur dataset for (a) SIRT with the initial gray level estimate, (b) SIRT with the corrected gray level estimate, (c) PDART with the initial gray level estimate, and (d) PDART after six correction steps.

For the segmented SIRT reconstruction, the reconstructed residual error (Fig. 9a) shows that the gray level for the bone is underestimated. Corrections for the gray levels were computed by applying (3) to the reconstructed residual error of Fig. 9a. With corrected gray levels for the segmented reconstruction, the reconstructed residual error (Fig. 9b) does not show a systematic error anymore, in the sense that the bone now contains both positive and negative errors. For the initial PDART reconstruction (Fig. 8c), the reconstructed residual error (Fig. 9c) also shows that the gray level for the bone is underestimated. Additionally, Fig. 9c suggests that the bone structures are too wide, since they have a region of negative errors at their edges. From Fig. 9c, the gray levels can again be corrected using (3). For this particular algorithm, it makes sense to redo the reconstruction with the corrected gray levels, since those gray levels are an input of the algorithm. Fig. 8d shows the resulting PDART reconstruction after six such correction steps.

4. Discussion

In tomography, most techniques that are used for segmentation and segmentation evaluation do no exploit the projection data. There are a few algorithms that do exploit this information during segmentation [17,18], however, they compute a quality measure that is a single number (the *projection distance* in [17] and the *segmentation inconsistency* in [18]). In contrast, the reconstructed residual error is a spatial map of the segmentation quality. This allows studying local variations of the error, which is not possible using a single number. Moreover, it allows evaluating any given segmentation, including those that have been computed through a technique that is not tomography-aware.

The reconstructed residual error can exploit the projection data because the artifacts of the reconstruction, which are typically largely in the null space of the system matrix (Figs. 3j and 4c), are largely in the row space after the segmentation step (Figs. 3b and 4e).

Computing the exact row space component of the segmentation artifacts is infeasible in practice, due to computational constraints and noise. However, a regular reconstruction algorithm such as SIRT can be substituted without much degradation of the results (compare Figs. 3c and 5a with Figs. 3b and 4e, respectively). This allows using the reconstructed residual error for practical datasets.

Using the reconstructed residual error, it seems possible to detect even small errors in the gray levels (Fig. 6b and e), as is evident from Table 2. Larger errors in the gray levels are visible too (Fig. 9a and b), but more important in this case is that structural problems in the original reconstruction are also revealed (Fig. 9c).

The results for experimental micro-CT data also illustrate another type of analysis that is enabled by the proposed method. After correction of the initial gray levels of the segmented SIRT reconstruction (Fig. 8b), the reconstructed residual error (Fig. 9b) suggests an uneven density of the bone. The conclusion might be that this effect is real, or that it is due to the scanning process. However, when the PDART algorithm is used (Fig. 8d), the uneven density almost completely disappears (Fig. 9d). Hence, this segmentation *is* compatible with the bone material having a homogeneous density. Given the relatively small differences between the segmented images (Fig. 8b and d), it is not obvious how to choose between them without access to the reconstructed residual error.

5. Conclusions

We have introduced the reconstructed residual error, as a way to evaluate the segmentation quality of a reconstructed image in cases where the segmentation is based on the density of the scanned object. We have used the Moore–Penrose pseudoinverse as a mathematical model for investigating the properties of the technique, and then generalized this approach to practical reconstruction algorithms.

The reconstructed residual error provides an accurate map of the errors in a segmented tomogram. Visually, this map can be used to study the distribution of the errors. Numerically, it can be used to improve the gray level estimates. The method has also been applied successfully to experimental datasets.

The necessary tools to implement the technique in practice are a forward projector and a reconstruction algorithm. The computational cost of the algorithm is modest, since only a single forward projection and a single reconstruction are needed. These aspects make the adoption of the method quite feasible in practice.

Acknowledgments

References

[1] T.M. Buzug, Computed Tomography: From Photon Statistics to Modern Cone-Beam CT, Springer-Verlag, Berlin/Heidelberg, 2008.
[2] R.M. Haralick, L.G. Shapiro, Image segmentation techniques, Comput. Vis. Graph. 29 (1) (1985) 100–132.
[3] N.R. Pal, S.K. Pal, A review on image segmentation techniques, Pattern Recogn. 26 (9) (1993) 1277–1294.
[4] D.L. Pham, C. Xu, J.L. Prince, Current methods in medical image segmentation, Annu. Rev. Biomed. Eng. 2 (2000) 315–337.
[5] Y.J. Zhang, A survey on evaluation methods for image segmentation, Pattern Recogn. 29 (8) (1996) 1335–1346.
[6] H. Zhang, J.E. Fritts, S.A. Goldman, Image segmentation evaluation: a survey of unsupervised methods, Comput. Vis. Image Underst. 110 (2) (2008) 260–280.
[7] P.L. Correia, F. Pereira, Stand-alone objective segmentation quality evaluation, EURASIP J. Appl. Signal Process. 2002 (1) (2002) 389.
[8] G. Strang, Introduction to Linear Algebra, Wellesley-Cambridge Press, Wellesley, MA, USA, 2009.
[9] D.A. Langan, J.W. Modestino, J. Zhang, Cluster validation for unsupervised stochastic model-based image segmentation, IEEE Trans. Image Process. 7 (2) (1998) 180–195.
[10] M. Beister, D. Kolditz, W.A. Kalender, Iterative reconstruction methods in X-ray CT, Phys. Med. 28 (2012) 94–108.
[11] K.J. Batenburg, J. Sijbers, DART: a practical reconstruction algorithm for discrete tomography, IEEE Trans. Image Process. 20 (9) (2011) 2542–2553.
[12] Y. Saad, Iterative Methods for Sparse Linear Systems, SIAM, Philadelphia, PA, USA, 2003.
[13] P. Gilbert, Iterative methods for the three-dimensional reconstruction of an object from projections, J. Theor. Biol. 36 (1) (1972) 105–117.
[14] J. Gregor, T. Benson, Computational analysis and improvement of SIRT, IEEE Trans. Med. Imag. 27 (7) (2008) 918–924.
[15] T. Roelandts, K.J. Batenburg, E. Biermans, C. Kübel, S. Bals, J. Sijbers, Accurate segmentation of dense nanoparticles by partially discrete electron tomography, Ultramicroscopy 114 (2012) 96–105.
[16] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst., Man, Cybernet. 9 (1) (1979) 62–66.
[17] K.J. Batenburg, J. Sijbers, Optimal threshold selection for tomogram segmentation by projection distance minimization, IEEE Trans. Med. Imag. 28 (5) (2009) 676–686.
[18] W. van Aarle, K.J. Batenburg, J. Sijbers, Optimal threshold selection for segmentation of dense homogeneous objects in tomographic reconstructions, IEEE Trans. Med. Imag. 30 (4) (2011) 980–989.