

Efficient low-dose CT artifact mitigation using an artifact-matched prior scan

Wei Xu^{a)} and Klaus Mueller^{b)}

Visual Analytics and Imaging Lab, Center of Visual Computing, Computer Science Department, Stony Brook University, Stony Brook, New York 11794-4400

(Received 10 November 2011; revised 5 June 2012; accepted for publication 12 June 2012; published 19 July 2012)

Purpose: Low-dose CT has attracted increasing attention due to growing concerns about radiation exposure in medical scans. However, the frugal use of x-ray radiation inevitably reduces the quality of the CT images, introducing artifacts such as noise and streaks which make the reconstructed images difficult to read in clinical routine. For follow-up CT exams a prior scan is often available. It typically contains the same anatomical structures, just somewhat deformed and not aligned. This work describes a two-step technique that utilizes this prior scan to achieve high-quality low-dose CT imaging, overcoming difficulties arising from noise artifacts and misalignment. We specifically focus on reducing the dose by lowering the number of projections. This gives rise to severe streak artifacts which possibly lower the readability of CT images to a larger extent than the fine-grained noise that results from lowering the mA or kV settings.

Methods: A common approach is to apply image filtering to reduce the noise artifacts. These techniques typically utilize pixel neighborhoods in the degraded image to estimate the true value of a pixel at the center of this neighborhood. However, this can lead to poor results when the image is severely contaminated under very low low-dose situations. We propose a method that utilizes the nondegraded, clean prior to determine higher quality pixel statistics to form the pixel estimates, supported by the matching scheme of the non-local means filter. To make this matching reliable, a good registration of prior and low-dose image is required. For this, we employ a state-of-the-art registration method, called SIFT-flow, which can tolerate the high amount of streak noise. But even for properly registered images, using an artifact free prior for the matching yields inferior results. We hence describe a scheme that first constructs a tandem-prior with streak artifacts resembling those in the low-dose image, and then employs this image for the matching, but uses the corresponding high-quality prior to determine the pixel estimates.

Results: Two experimental studies are carried out, using a head phantom and a human lung with projections gathered via simulation. We assess the quality of the processed reconstruction with various metrics: mathematical and perceptual. We find that the quality that can be obtained with the artifact-matched prior-based scheme significantly exceeds that of all competing schemes. Even though the general prior-based approach is able to eliminate the streak artifacts, only the artifact-matched scheme can restore small detail and feature sharpness.

Conclusions: The reduced-projection low-dose image reconstruction algorithm we present outperforms traditional image restoration algorithms when a prior scan is available. Our method is quite efficient and as such it is well suited for fast-paced clinical applications such as image-assisted interventions, orthopedic alignment scans, and follow-ups. © 2012 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4736528>]

Key words: artifact-matched prior, artifact mitigation, low-dose CT, nonlocal means, reference-based

I. INTRODUCTION

Computed x-ray tomography has revolutionized modern medicine and thanks to the rapid growth in scanner technology the gamut of its applications has risen at an enormous rate. In this process, buoyed by the excitement of possibilities little attention was paid to the radiation dose administered to the patient. Scans with ever-improving spatial and temporal resolutions were conducted on a routine basis and the associated CT reconstruction algorithms had the luxury of an abundance of data collected at each exam. It was only recently that the sobering results of long-term studies on the

adverse radiation effects of CT imaging have dampened these developments.^{1,2} Due to these studies, the harmful effects of x-ray radiation in CT scans have become publicly heard, threatening the future of this modality. To counter these concerns, campaigns such as ImageGently (Ref. 29) and ImageWisely (Ref. 30) have been initiated that promote the optimization of the radiation dose used in both pediatric and adult medical imaging.

To reduce the radiation dose subjected to the patient one can: (1) lower the number of scans, (2) lower the number of x-ray projections per scan, and (3) lower the energy settings of the x-ray tube (kV, mA) per projection image. The first

measure, i.e., reducing the number of scans, is often left at the discretion of the treating physician. The latter two options are highly detrimental to image quality, resulting in images with significant noise artifacts. They greatly challenge the conventional CT reconstruction algorithms based on analytical formulations rooted in the inverse radon transform,³ and these shortcomings have recently invigorated research efforts towards methods that seek alternatives to these conventional schemes.

A popular approach to this end has been to enforce data fidelity and image quality as a joint optimization problem and solve these two parts in an iterative round-robin fashion. Data fidelity can be assured by ways of any CT reconstruction algorithm, iterative or analytical, but most use the former. The reduction of noise artifacts, on the other hand, can be posed as an image denoising problem. Many approaches use the method of total variation minimization (TVM) (Ref. 4) for this task since it is often part of general compressive sensing formulations⁵ that were originally prescribed to deal with sparse data. For CT reconstruction, a number of sophisticated schemes have been developed that adapt the various parameters used in the process, such as ASD-POCS (Ref. 6) and soft-threshold filtering.⁷

In this work, we have attempted to devise a framework that executes the data fidelity step and the image quality step each exactly once. It is hence of lower computational complexity than the present schemes which perform these steps iteratively. We achieve this by making creative use of an artifact-free prior—constituted by an existing regular-dose scan of the patient. Such a clean prior scan is frequently available. For example, it may be a regular-dose first scan acquired before a low-dose follow-up scan or it may be a regular-dose diagnostic scan preceding a low-dose setup scan for a surgical intervention such as orthopedic spine fixation, among other scenarios.

In the present work, we focus on the second form of low-dose CT, i.e., reducing the number of x-ray projections per scan. The first step of our framework uses filtered backprojection (FBP) to reconstruct an image with significant streak artifacts which result from this low number of projections. The use of FBP to provide a quick first estimate is a common strategy. Unlike the first step of an iterative scheme, such as ART (Ref. 8) and its derivatives,^{9,10} FBP typically reconstructs all image features at good fidelity but the high image noise makes them difficult to read. Common approaches then follow FBP by an iterative pipeline for denoising. Our second step, on the other hand, uses a single prior-based image restoration that eliminates the noise and so provides the desired viewing experience. In this step, we first register/align our prior with the FBP-estimate using an established multiscale feature registration algorithm, i.e., SIFT flow.¹¹ Following, we simulate the low-dose streak artifacts of the FBP-estimate in this registered clean prior. Finally, for each pixel in the target image we use a neighborhood similarity metric to determine the best matches in the contaminated prior and then replace it using the corresponding pixels in the clean prior.

Using existing scans to support regularization is not new. Kelm *et al.*¹² describe an approach that reconstructs volumes

at two different thicknesses, using the same acquired projection data. They reconstruct the thicker slices from bin-averaged projections which increases SNR, while the thinner (and noisier) slices are reconstructed from the original projection data. Since registration is implicit, it is relatively straightforward to use the thicker slices for neighborhood-based denoising of the thinner slices. In contrast, our method applies to settings in which the reference images are not necessarily acquired simultaneously. Yu *et al.*¹³ present the method previous scan-regularized reconstruction (PSRR). It replaces regions that are unchanged in a low-dose CT reconstruction with their direct embodiments in a normal-dose CT reconstruction and uses a nonlinear diffusion approach for denoising in the remaining regions. This approach requires an effective strategy for feature recognition, which the authors accomplish via registration.

The approach most similar to ours is that of Ma *et al.*¹⁴ They also use a registration algorithm—a combination of rigid principal component analysis (PCA) and nonrigid mutual information optimization¹⁵—for rough alignment. They then use a neighborhood-mechanism to locate suitable replacement candidates in a prior scan of the patient. Our work differs from theirs in the following important ways. First, while Ma *et al.* restore images reconstructed from projections generated at reduced mA settings, we treat artifacts caused by the reduction of projections. The resulting streak artifacts are much more severe and irregular than the random noise-artifacts caused by low-mA imaging. Second, instead of using a clean prior for matching we use a prior with simulated artifacts. We find that this affords much better accuracy, as we will prove and demonstrate. A preliminary version of the framework we describe here has been presented in Ref. 16, which predates the work by Ma *et al.* slightly.

Our paper is organized as follows. Section II presents the algorithms we have studied, Sec. III presents results, and Sec. IV ends with conclusions and points to future work.

II. METHODS AND MATERIALS

To locate good pixel matches in the prior we use the similarity measures also employed by nonlocal means (NLM) filtering.¹⁷ NLM filtering can be seen as a generalization of Gaussian smoothing. It looks for structurally similar pixel neighborhoods in the smoothing site's proximity and includes them into the Gaussian filter statistics. This leads to a more robust estimate of the true pixel value and consequently to improved image restoration/denoising results. We have recently informally compared NLM filtering with TVM for low-dose imaging tasks and our results have been quite encouraging.¹⁸ In this current work, we do not employ NLM-filtering in a conventional way as a nonlocal extension of Gaussian filtering. Rather, we only retain NLM's mechanism for similarity-based pixel neighborhood matching in the artifact-matched prior.

II.A. Standard NLM filtering

The NLM algorithm was proposed by Buades *et al.*¹⁷ for image denoising. It takes advantage of the high degree of

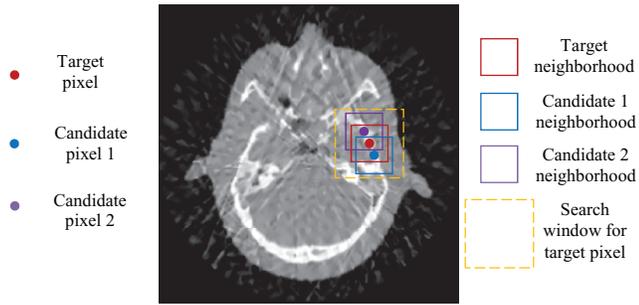


FIG. 1. Illustration of conventional NLM for denoising.

redundancy that typically exists in an image. Given a target pixel subject to denoising, it defines a small Gaussian-weighted region around it, called a *patch*. It then searches the entire image for similar patches and accumulated them weighted by their degree of similarity. In practice, only a local neighborhood around the target pixel is searched, called *search window*. This helps performance but it also better tolerates nonstationary noise processes. Further, since we first register the prior to the target image, we do not require large search windows in any case. More formally, the updated value p'_x of a target pixel is computed as

$$p'_x = \frac{\sum_{y \in W_x} \exp\left(-\sum_{t \in P} G_a(t) |p_{x+t} - p_{y+t}|^2 / h^2\right) \cdot p_y}{\sum_{y \in W_x} \exp\left(-\sum_{t \in P} G_a(t) |p_{x+t} - p_{y+t}|^2 / h^2\right)}. \quad (1)$$

Here, x is the location of the target pixel and the y are the locations of the candidate pixels, with values p_y . W_x is the search window around x , and P is the patch size of each pixel. The patch similarity is measured by the Gaussian weighted L_2 distance between two patch vectors with t representing the index within a patch and G_a being a Gaussian kernel with standard deviation a . The exponential function converts these distances to weights, determined by a parameter h which controls the overall smoothness of the filtering. Larger values of h will result in more smoothing.

Figure 1 shows an illustration of this process when NLM is used in a conventional way to denoise a reconstruction with severe streak artifacts. In this particular case we used FBP to reconstruct a GE head phantom from 45 fan-beam projections acquired over 360° . The illustration shows the search window, the target pixel in the center and two candidate pixels with similar neighborhoods as the target pixel. Figure 2(a) shows the NLM-filtered result, while Fig. 2(b) shows the results obtained with TVM. We can observe that both of these filters provide some amount of improvement over the original image shown in Fig. 5(a). It is likely (see, for example, Ref. 19) that repeating the fidelity and denoising steps several times would do substantially better, but since we have aimed for a nonrepeating approach—one in which a prior is available to aid in the denoising—we focus on a single step scheme.

II.B. Registration using the SIFT-flow algorithm

A crucial element in our prior-assisted framework is proper registration since without it the possibility for mismatches can

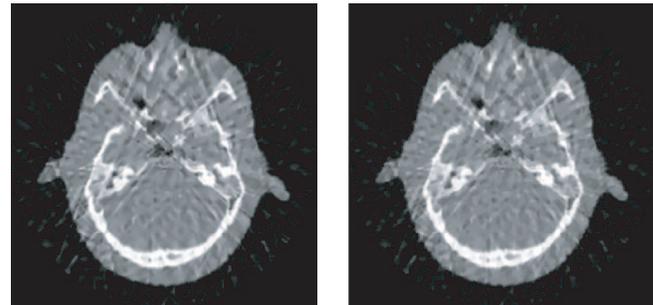


FIG. 2. Filtering results obtained with NLM and TVM.

be high. Nevertheless, our use of the NLM-based matching mechanism relaxes the need for tight and laborious registration of the prior image as it performs the fine registration on the fly via its search mechanism. As such it is less sensitive to spatial distortions than the PSRR approach. For registration we have made use of the SIFT-flow algorithm, recently published by Liu *et al.*¹¹ This algorithm originates from the optical-flow algorithm which produces dense, pixel-to-pixel correspondences between two images. It extends the matching from raw pixels to SIFT feature descriptors.²⁰ A scale-invariant feature transform (SIFT) feature descriptor captures the histogram of gradient orientations in a local neighborhood at a given scale. It is well suited to characterize salient local and transform-invariant image structures and at the same time encode contextual information. SIFT-flow has been specifically designed for scene matching, where objects share similar scene characteristics but may have different appearances and locate at different places. This is the case in the registration of the prior to the current scan. They are certainly from the same person, i.e., they share the scene characteristics, but they will likely have different SNR and undergone distortions of the features.

The implementation of SIFT-flow has two parts: (i) generate dense SIFT features where each pixel has a 128-dimensional SIFT vector, and (ii) find the correspondence of these SIFT features via discrete optimization on the image lattice to obtain the displacement field for alignment. For the first part, Fig. 3 shows a typical SIFT feature descriptor summarizing the gradient orientations in a 16^2 pixel area (as plotted inside the red square). The gradients (shown as blue arrows) are Gaussian-smoothed according to their distance to the area center. This area is partitioned into 4^2 blocks, each of size 4^2 pixels (shown as green squares). The gradient orientations are then accumulated in each block to 8 orientation bins and are weighted by their gradient magnitudes. There are a total of 16 8-bin orientation histograms (with each red arrow representing one bin). Thus the dimension of a SIFT vector is $4 \times 4 \times 8 = 128$ over a $16 \times 16 = 256$ area. For the second part, to estimate flow, the energy function for SIFT flow is defined as below

$$E(f(x, y)) = \sum_{(x, y)} \min(\|s_1(x, y) - s_2(x + f_x(x, y), y + f_y(x, y))\|_1, a)$$

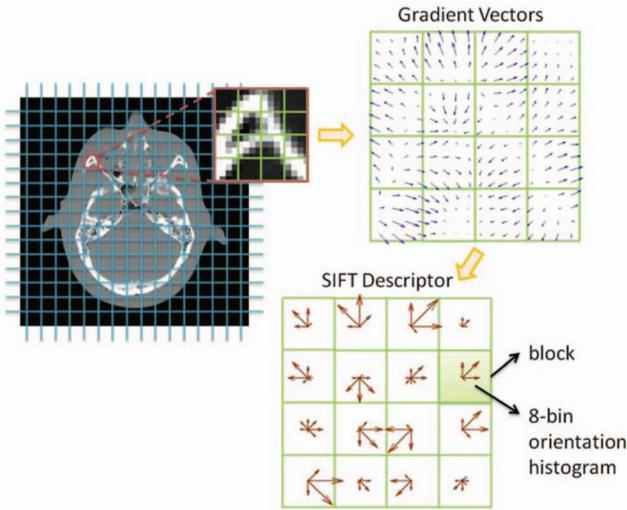


FIG. 3. Illustration of SIFT descriptor summarizing edge orientations over 16×16 pixel area.

$$\begin{aligned}
 & + \sum_{(x,y)} b(|f_x(x,y)| + |f_y(x,y)|) \\
 & + \sum_{(x',y') \in W(x,y)} (\min(c|f_x(x,y) - f_x(x',y')|, d) \\
 & + \min(c|f_y(x,y) - f_y(x',y')|, d)), \quad (2)
 \end{aligned}$$

where (x, y) and (x', y') are pixel locations, s_1 and s_2 are SIFT descriptors, f is the displacement function with f_x in x -direction and f_y in y -direction, W is the pixel neighborhood and a, b, c, d are four thresholds (with default settings). This function is designed according to three constraints: (1) the matched pixel should have similar SIFT descriptors; (2) the displacement should be as small as possible; and (3) adjacent pixels should have similar displacements to maintain flow smoothness. This discrete displacement function can be estimated by optimizing the energy function with a belief propagation algorithm.¹¹ Its time complexity is $O(h^2 \log h)$ where h is the width (height) of the image. Before registration we smooth the images with a 7×7 Gaussian filter (standard deviation = 3). This yields more stable results. Using the MATLAB implementation obtained from the author's website²¹ it took less than 1 min for one registration operation of two 256^2 CT scans on a quad-core Dell XPS 2.66 GHz PC with 8GB of memory. In their paper, Liu *et al.*¹¹ also point out that a GPU implementation of their belief propagation algorithm could yield a further (up to) 50-time speedup which would bring the time required for the registration down to seconds.

II.C. Reference-based NLM (R-NLM) filtering

With the NLM-mechanism still being employed for the matching, we call our approach reference-based NLM (R-NLM) filtering since it uses the prior image as a reference to guide the filtering. Figure 4 provides an illustration of this process, now using a noticeable smaller search window than in the regular NLM-case. The larger the search window the

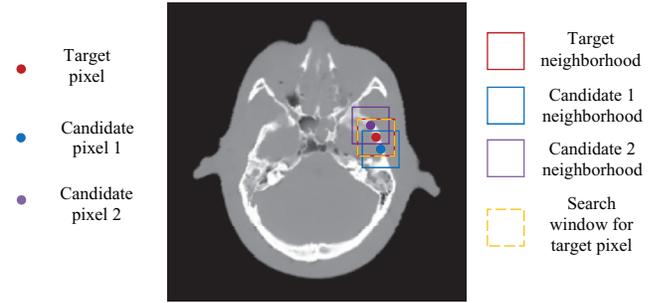


FIG. 4. Illustration of R-NLM with the registered clean prior used for both matching and retrieval.

more distortion-tolerant the algorithm becomes, but the potential for lost detail and oversmoothing also rises. In experiments we found a size of 7×7 pixels for both search window and patches to represent a good compromise.

Our R-NLM algorithm first uses SIFT-flow to align the prior with the current scan, call it *target scan*. Following, it visits every pixel in the target scan, places the search window in the same location in the prior scan, and uses the NLM-algorithm to determine the update. Figure 5 presents some results we have obtained with our R-NLM algorithm, to motivate a further extension discussed in Sec. II.D. Figure 5(f) shows a regular-dose reconstruction obtained with 360 projections, while Fig. 5(a) shows a low-dose FBP reconstruction obtained from the same data but only 45 projections—about 1/8 of the dose. Figure 5(b) shows the prior. Since this was a head phantom that could not be warped mechanically, we performed a digital warp—in this case a twirl distortion around the center of the image. This deformation field is shown in Fig. 5(h). Figure 5(c) shows the registered prior and Fig. 5(d) shows the result obtained with R-NLM filtering. It is clearly better than the NLM and TVM filtered results (see Fig. 2). This of course is a comparison that is only partially fair because R-NLM had access to a clean prior, while the others did not. This procedure changes Eq. (1) into the following:

$$p'_x = \frac{\sum_{y \in W_x} \exp\left(-\sum_{t \in P} G_a(t) |p_{x+t} - p_{y+t}^{crp}|^2 / h^2\right) \cdot p_y^{crp}}{\sum_{y \in W_x} \exp\left(-\sum_{t \in P} G_a(t) |p_{x+t} - p_{y+t}^{crp}|^2 / h^2\right)}. \quad (3)$$

Here, the superscript *crp* indicates that the pixels originate from the clean registered prior and not from the target. However, when comparing this result with that obtained at regular dose [Fig. 5(f)], we still observe some amount of blurring in the image. Edges in general appear less defined, and small features are also weakened or completely suppressed. For the latter, compare for example the intricate detail in the center of the image, to the left of the pincushion-shaped dark structure, which is barely visible in the R-NLM result. In Sec. II.D, we describe an advanced scheme that overcomes these problems.

Finally, in the event that no reliable update can be found in the prior for a given target pixel, we fall back to conventional NLM-filtering using the target image. We identify this situation by a low sum of weights in Eq. (3). In our experiments,

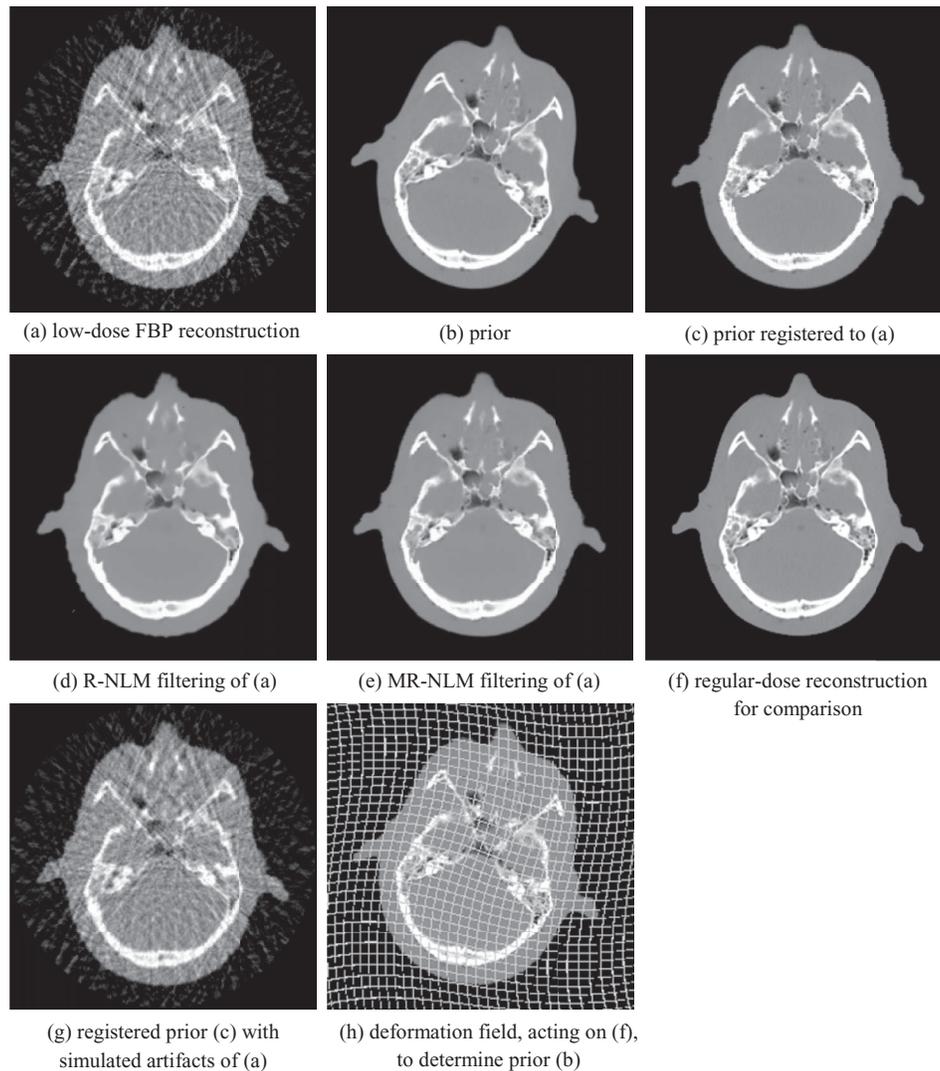


FIG. 5. Results-supported illustration of reference-based NLM (R-NLM) and matched reference-based NLM (MR-NLM).

we have used a threshold of 0.001. We use this criterion both for R-NLM and for the advanced scheme described next.

II.D. Matched reference-based NLM (MR-NLM) filtering

Since the features are generally weakened at a scale less than the size of the NLM search window, we cannot blame the registration algorithm for these shortcomings. Rather, it is the quality of the NLM-matching that is at the heart of the problem. Consider the NLM-distance function of Eqs. (1) and (3) used to determine the quality of a match for a specific candidate neighborhood (or patch) P ,

$$\sum_{t \in P} G_a(t) |p_{x+t} - p_{y+t}|^2. \quad (4)$$

Here, $|p_{x+t} - p_{y+t}|$ is the Euclidian distance of a corresponding pair of pixels parameterized by patch index t . The sum of these distances determines the weight that the patch P plays in determining the value of the target pixel, and thus it is the patch's structural similarity that is decisive for the scaling of its contribution. While Eqs. (1) and (3) also have a

parameter h for scaling, it is a global parameter that scales all patches at the same weight. The difficulties we encounter with R-NLM cannot be solved just by adjusting the factor h , as we will demonstrate in Sec. II.E. Just as the best registration is achieved when the two scenes are similar in appearance, the best NLM-match is obtained when target and prior have a similar appearance. This is not the case when pairing a clean prior and a low-dose reconstruction with severe streak artifacts. Hence, we require a method that transforms the clean registered prior image into an image that bears similar artifacts as the target image. We can achieve this by first simulating projections from the registered prior and then reconstructing it under the same conditions as the target image, i.e., with a lower number of projections and at the same viewing geometry as the target. This gives rise to Fig. 5(g)—which as we observe looks fairly close to the target image subject to denoising [see Fig. 5(a)].

The MR-NLM reconstruction procedure is illustrated in form of pseudocode in Fig. 6. After registering the prior with the low-dose target, a degraded registered prior is created by simulating the low-dose artifact also present in the target. The

Input:
 Low-dose scan L , normal-dose prior scan N , low-dose degradation D ;

Preprocessing:

1. Register N to L using SIFT-flow and obtain N_R :
 $N_R \leftarrow \text{SIFT_Flow_Registration}(N, L)$;
2. Generate projections P of N_R with the same low-dose degradation D as the input:
 $P \leftarrow \text{Forward_Projection}(N_R, D)$;
3. Generate the degraded version of $N_R - N_{DR}$ using FBP:
 $N_{DR} \leftarrow \text{FBP}(P, D)$;

Filtering:
 Apply MR-NLM filtering to L with $\langle N_R, N_{DR} \rangle$ and return the denoised filtered result L_F :
 $L_F \leftarrow \text{MR-NLM}(N_R, N_{DR}, L)$;

FIG. 6. Pseudocode of the matched reference-based NLM (MR-NLM).

procedure then uses this degraded registered prior for NLM-matching, but copies the corresponding candidate pixels in the clean registered prior to the weighted sum. The resulting equation is, modifying Eq. (3),

$$p'_x = \frac{\sum_{y \in W_x} \exp\left(-\sum_{t \in P} G_a(t) |p_{x+t} - p_{y+t}^{drp}|^2 / h^2\right) \cdot p_y^{crp}}{\sum_{y \in W_x} \exp\left(-\sum_{t \in P} G_a(t) |p_{x+t} - p_{y+t}^{drp}|^2 / h^2\right)}. \quad (5)$$

Here the subscript crp denotes the clean registered prior, as before, while the subscript drp denotes the degraded registered prior.

Figure 5(e) shows the result we obtained for our test example. We observe that the edges are now overall significantly

sharper, small features are better visible, and we also see that the intricate detail in the center of the image, to the left of the pincushion-shaped dark structure is also clearly restored.

II.E. Comparing NLM, R-NLM, and MR-NLM

Figure 7 compares the three different schemes, using a case study at “microscopic” detail. In Fig. 7(a), we show a clean registered prior—the reference. For the matter of this discussion, we shall focus on the 7×7 image cutout—equivalent to a NLM search window—within the black box in the lower right half of this image. In Fig. 7(b), we illustrate the data flow and operations of all three schemes using the cutout as an example. In this schematic, the cutout labeled “clean reference” is a copy of the black-boxed region, while the “target” is the corresponding cutout in the low-dose scan which is much degraded. As discussed, the MR-NLM procedure first simulates the low-dose artifacts in the clean reference producing the “degraded reference.” It then uses this image for NLM-matching, but retains the corresponding pixels in the clean reference to update the target, yielding the cutout labeled “MR-NLM.” On the other hand, the R-NLM procedure (dotted lines) uses the clean reference for matching and updates the target directly, giving rise to the cutout labeled “R-NLM.” Finally, the conventional NLM procedure uses the target for both match and update, producing the cutout labeled

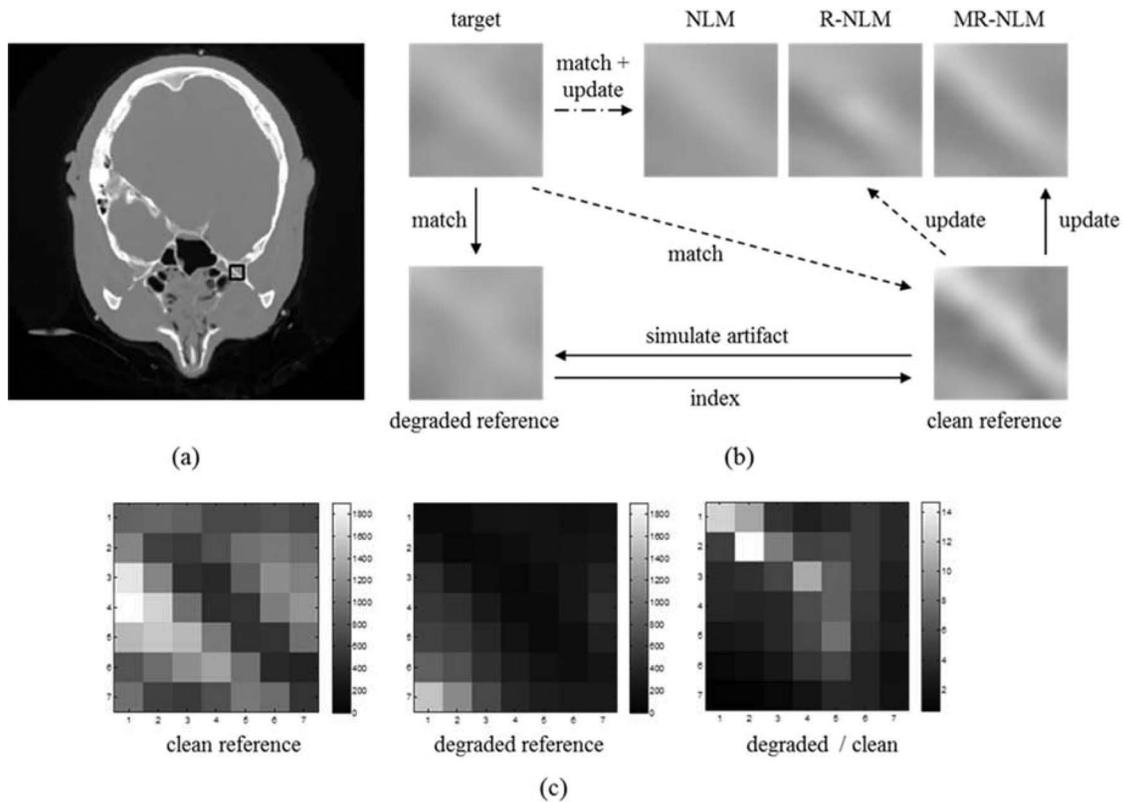


FIG. 7. Comparing the NLM, R-NLM, and MR-NLM filtering schemes in terms of their effect on a 7×7 image region equivalent to the size of a search window. (a) Registered clean prior with black box in lower right image region indicating the region studied. (b) The three pipelines illustrated by example. (c) The map of pixel contributions for this search window. Each pixel is associated with a 7×7 patch centered on it.

“NLM.” The row of result cutouts demonstrates an increasing growth in quality from left to right. While the NLM cutout is quite similar to the low-dose target subject to denoising, the R-NLM cutout has somewhat sharper detail, in particular in the center. Finally, the MR-NLM cutout has the most pronounced sharpness—not quite as strong as the clean reference but fairly close.

Further insight can be obtained from visualizing the distance map for the NLM search window coinciding with the studied image cutout. Note that this search window will only resolve the value for the pixel in the center of the cutout. This distance map is used for the matching—see Eq. (4). Plotted in Fig. 7(c) are the corresponding maps for the clean reference used in R-NLM and for the degraded reference used in MR-NLM, respectively. We can easily see that the distances in the latter map are much closer than those in the former which confirms the better correspondence. The third map, labeled “degraded/clean” shows the ratio of the two maps. We clearly see that this ratio is not constant across the patch and thus a simple boosting of the h -parameter in the NLM equations would not be able to rectify this situation.

As for the efficiency of the three algorithms, the computational complexity is $O(NWP)$ where N , W , and P represent image size, search window size and patch size, respectively. However, in a GPU implementation, due to the high pixel independence and therefore potential parallelism, the speed could be greatly increased. As determined in Ref. 22 for the 2D case, it only takes 18 ms to denoise a 512^2 image with a 11^2 search window and a 7^2 patch size. MR-NLM (R-NLM) incurs a small additional overhead for reading from two (one) other image(s) and for checking if falling back to conventional NLM-filtering is more appropriate. Overall, the entire algorithm, including FBP reconstruction (see Ref. 23), SIFT-flow registration (see Sec. II.B), and MR-NLM or R-NLM filtering, would most likely take on the order of seconds when accelerated on a high-performance GPU.

II.F. Assessing image quality

To evaluate the quality performance of the various reconstruction schemes we have employed two groups of metrics. The first group encompasses the traditional RMS (root mean square) and CC (correlation coefficient) measures defined as follows:

$$\begin{aligned} \text{RMS} &= \sqrt{\frac{\sum_{i=1}^N (p_{l,i} - p_{r,i})^2}{N}} \\ \text{CC} &= \frac{\sum_{i=1}^N (p_{l,i} - \mu_l)(p_{r,i} - \mu_r)}{\sqrt{\sum_{i=1}^N (p_{l,i} - \mu_l)^2 \sum_{i=1}^N (p_{r,i} - \mu_r)^2}}. \end{aligned} \quad (6)$$

In these metrics, the $p_{l,i}$ are the pixels in the low-dose reconstruction and the $p_{r,i}$ are the pixels in the corresponding regular-dose reconstruction, in our case constituted by the originally obtained scan image. (We note that in our exper-

iments the prior is created by nonlinearly distorting the original scan, the low-dose image is created in alignment with the original scan, and the registration brings the prior back into *approximate* alignment with the original scan. Thus, the most appropriate gold standard is the originally obtained scan). The μ_l and μ_r are the averages of the low and regular dose images l and r , respectively, and N is the total number of pixels.

The advantage of these metrics is that they are easy to compute and have clear physical meanings. However, they reveal only little about the perceptual impact certain image differences may have. The RMS metric computes the pointwise errors and pools them across the entire image—this ignores any spatial coherence and so cannot gauge the differences in structure and contrast that may exist in local pixel neighborhoods. On the other hand, while CC does provide a statistical measure of image differences, it computes it at a global scale and it also considers only pixel intensities which are far less perceptually salient than local contrasts and edges.

As an attempt to better account for human perception when determining image quality, we have employed metrics that specifically gauge the preservation of perceptually salient information, which we define as image content to which the human visual system is most sensitive to.

The first such metric is E-CC, as defined in our earlier work.²⁴ It is identical to CC but operates on the edge-filtered images which we obtain using a Sobel mask. E-CC is still a global operator but it considers more perceptually salient low-level image features, i.e., edges that define the boundaries of the reconstructed objects.

Another metric we employ is the structural similarity index (SSIM) devised by Wang *et al.*²⁵ The SSIM is an enhancement of the universal image quality index (UQI) (Ref. 26) also recently used by Bian *et al.*¹⁹ Both UQI and SSIM combine the differences in mean intensity, contrast, and structure into a single quality figure. The SSIM is computed for each image pixel at position x_j over a sliding small image window—we use an 11×11 mask—and then combined into a pooled index $\text{SSIM}_{\text{pooled}}$ by averaging the individual SSIM measurements

$$\begin{aligned} \text{SSIM} &= \left(\frac{2\mu_l\mu_r + c_1}{\mu_l^2 + \mu_r^2 + c_1} \right) \left(\frac{2\sigma_l\sigma_r + c_2}{\sigma_l^2 + \sigma_r^2 + c_2} \right) \left(\frac{\sigma_{lr} + c_3}{\sigma_l\sigma_r + c_3} \right) \\ \text{SSIM}_{\text{pooled}} &= \frac{1}{N} \sum_{j=1}^N \text{SSIM}(x_j). \end{aligned} \quad (7)$$

Here, the subscripts l and r denote the low-dose and regular-dose images, respectively, and the μ_l and μ_r are the means of the pixels within these corresponding windows, while the σ_l and σ_r are their standard deviations and the σ_{lr} is their covariance. The constants c_1 , c_2 , c_3 are typically small (see Ref. 25) and prevent numerical instabilities when a denominator is close to zero—the UQI does not have these constants which can lead to wrong estimates when these adverse conditions are met. Finally, to avoid blocking artifacts Wang *et al.* recommend a Gaussian-weighting of the samples under a SSIM window. Since SSIM is the generally accepted name of the metric, we will use it throughout the paper but it is understood that we use its pooled version.

The first term in Eq. (7) is quite consistent with the just-noticeable intensity difference (JND) metric often used in perceptual quality studies. The second term compares the local contrasts that exist in the sliding window. Finally, the third term evaluates the structural similarity after the differences in means and contrasts have been accounted for. The SSIM is quite powerful—studies that ask human observers to rank images with identical scenes, but corrupted with different artifacts, in terms of quality show that these ranking correlate exceedingly well with the SSIM outcome. Furthermore, it is also interesting that in these studies all images had the same RMS error. Finally, large experiments²⁷ have shown that SSIM is particularly well suited to detect distortions caused by noise. It also detects spatially correlated noise, which in CT images could mimic false features.

III. RESULTS

We have run the algorithms described above on two datasets: a head phantom and a human lung. The head phantom is part of a body phantom scanned with a GE Light-Speed scanner. The human lung scan was obtained from the “Give a Scan” dataset collection.³¹ Specifically we used the first dataset series 2 of patient p0015 obtained with a GE LightSpeed16 scanner. For all examples, we used the original floating-point reconstructions for three purposes. First, they served as the basis for a high-quality projection simulation in fan-beam geometry (fan angle = 20°). We then picked a subset of these projections and reconstructed the reduced-projections low-dose imagery studied in this paper. Second, we also used them to generate the priors. For this, we applied various nonlinear distortions on them and subsequently registered them to the low-dose reconstructions. Third, they represented the gold standard for all numerical quality assessment via the various metrics described in Sec. II.F.

We begin with the head phantom already examined in Sec. II to illustrate the outcomes of the various algorithms. Table I and Fig. 8 compare the results obtained for MR-NLM, R-NLM, NLM, TVM, and no filtering, as gauged by the RMS, CC, E-CC, and SSIM error metrics. Table I also gives the settings for the various algorithm parameters which we manually

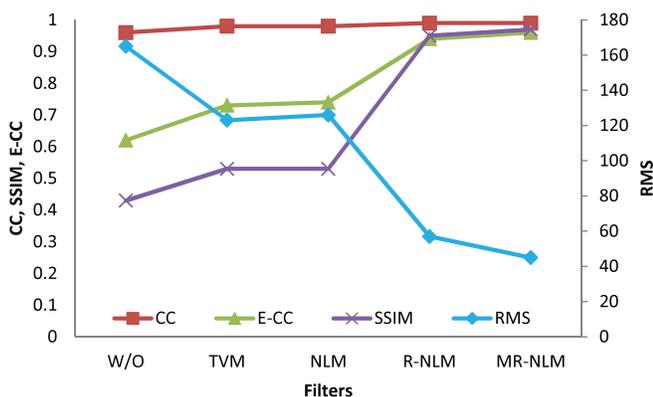


FIG. 8. Graphical comparison of the results obtained for the head phantom via various error metrics.

TABLE I. Numerical comparison of the results obtained for the head phantom via various metrics. The percentage figure for a metric measures the improvement with respect to the method to its immediate left. To the left of the % cell, above the scores, we list the optimal parameter setting for each algorithm which we obtained by manual tuning.

	W/O	TVM		NLM		R-NLM		MR-NLM	
	N/A	$\lambda = 30$	(%)	$h = 220$	(%)	$h = 200$	(%)	$h = 120$	(%)
RMS	165	123	25.4	126	2.4	57	54.8	45	21.0
CC	0.96	0.98	2.1	0.98	0	0.99	1.0	0.99	0
E-CC	0.62	0.73	17.7	0.74	1.4	0.94	27.0	0.96	2.1
SSIM	0.43	0.53	23.2	0.53	0	0.95	79.2	0.97	2.1

tuned for optimal performance. The first observation we make is that all metrics show similar trends (but we also observe that CC is much less sensitive to the changes in image quality). In general, for the CC, E-CC, and SSIM the maximum possible value is 1.0, while for the RMS error the optimal value is 0. In this particular experiment, all metrics reach their best values for the MR-NLM algorithm—around 0.97 for the perceptual metrics—and their worst values when no filtering is applied. It also appears that NLM and TVM reach quite similar scores for each metric, with a slight advantage for NLM. This can be verified by comparing the images (see Fig. 2) which look fairly similar. The improvement of R-NLM over NLM is significant for both RMS (54%) and the perceptual metrics (27% for E-CC and 79% for SSIM). The improvement for MR-NLM over R-NLM is another 21% for RMS, and 2% for the two perceptual metrics.

The absolute difference images presented in Fig. 9 show similar trends. Marked improvements can be observed for R-NLM over the prior-less schemes, and more moderately for MR-NLM over R-NLM. The improvement achieved by R-NLM is mainly in streak removal. MR-NLM, on the other hand, adds sharpness and detail definition which can be appreciated by the overall much smaller errors, in particular at edges and sharp corners.

In order to better explore the performance of the various algorithms at the local detail level we have conducted a ROI-based analysis (see Fig. 10). Figure 10(a) depicts the locations of four ROI regions and Fig. 10(b) plots the corresponding SSIM scores. We observe a 7%–8% improvement for R-NLM over NLM in all four ROIs and another 7%–8% in ROI 2, 3, and 4 for MR-NLM over R-NLM. ROI 1 contains a fairly structured and high-contrast feature which is less in need of

TABLE II. Numerical comparison of the results obtained for the human lung via various metrics.

	W/O	TVM		NLM		R-NLM		MR-NLM	
	N/A	$\lambda = 20$	(%)	$h = 260$	(%)	$h = 180$	(%)	$h = 120$	(%)
RMS	222	154	30.6	153	0.6	128	16.3	120	6.2
CC	0.92	0.96	4.3	0.96	0	0.97	1.0	0.98	1.0
E-CC	0.6	0.69	15.0	0.71	2.9	0.81	14.1	0.85	4.9
SSIM	0.49	0.61	24.5	0.61	0	0.66	8.2	0.69	4.6

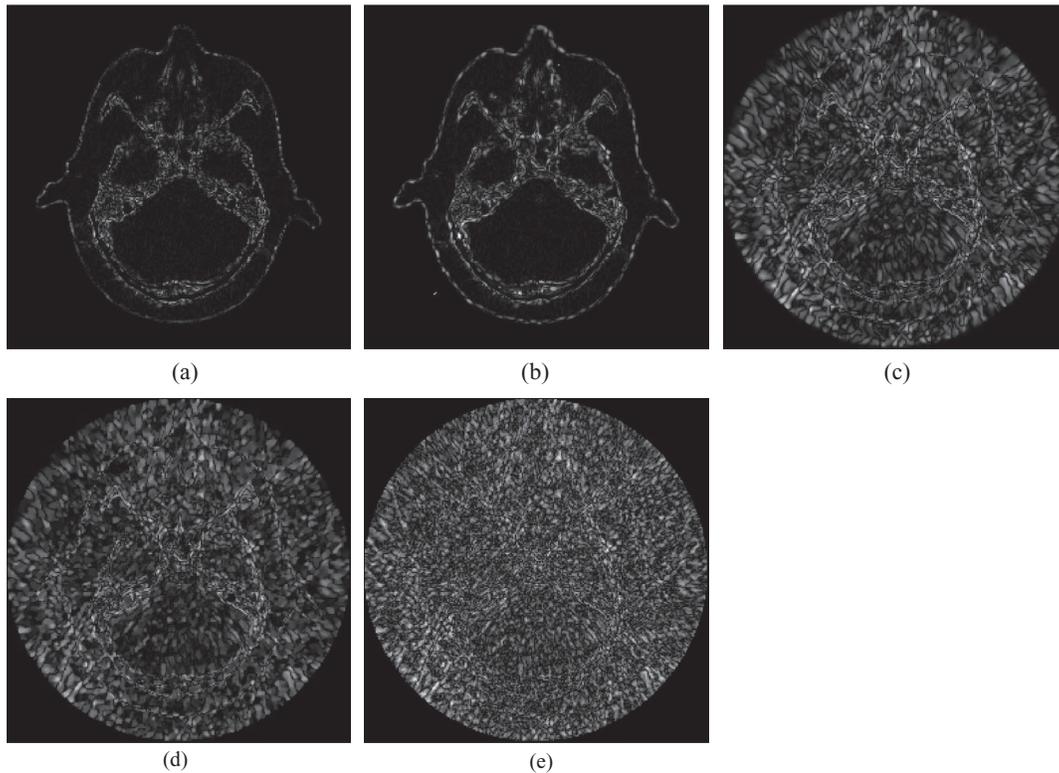


FIG. 9. Absolute difference images for the head phantom reconstructions. (a) MR-NLM; (b) R-NLM; (c) NLM; (d) TVM; and (e) without filtering.

the added fidelity of the MR-NLM scheme. Lastly, Fig. 11 presents the ROI study visually in form of cutout details, which best demonstrate the considerable benefits the MR-NLM restoration method provides. We observe that all ROIs show significantly more detail for MR-NLM, as opposed to R-NLM. In fact, the reconstructions are quite close to the ideal image. On the other hand, the differences of R-NLM vs the prior-less methods are also significant, but not as marked as for MR-NLM vs R-NLM.

Next, Fig. 12 show the same sequence of results obtained with the human lung, reconstructed from 90 projections over 360°. The distortion applied was a fisheye warp. Figure 13 presents difference images, and Table II lists and Fig. 14

plots overall evaluations with the various metrics. We make similar observations as for the head phantoms but note that in this test case the quantitative improvements for R-NLM and MR-NLM are more balanced and the MR-NLM/R-NLM gain is about double than that for the head phantom. Finally, Fig. 15 depicts ROI-definitions and the SSIM-scores for the studied restoration schemes. Figure 16 shows the cutout details. Again, we see that MR-NLM significantly improves the fidelity of small detail and in fact it is even able to restore some of the original CT image noise texture that was part of the prior.

Finally, Fig. 17 explores the effect of different search window sizes, for the human lung. The size needed mainly

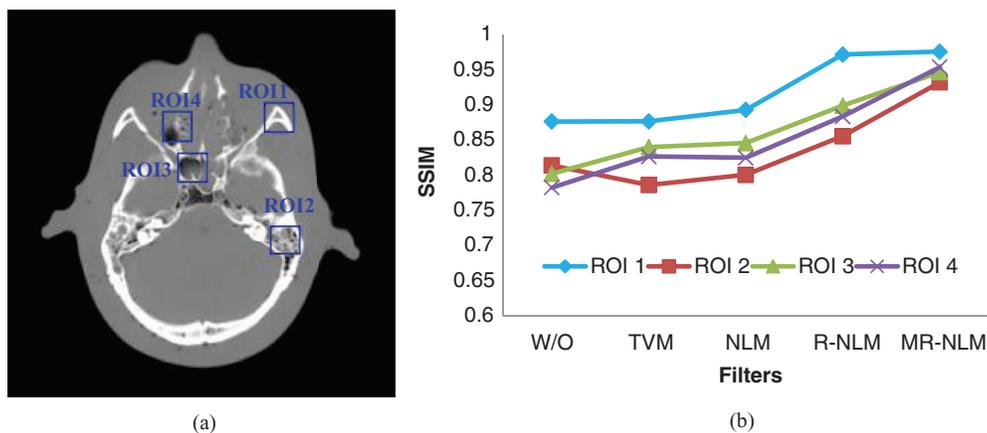


FIG. 10. ROI-based analysis for one slice of the head phantom. (a) ROI locations and (b) SSIM evaluation.

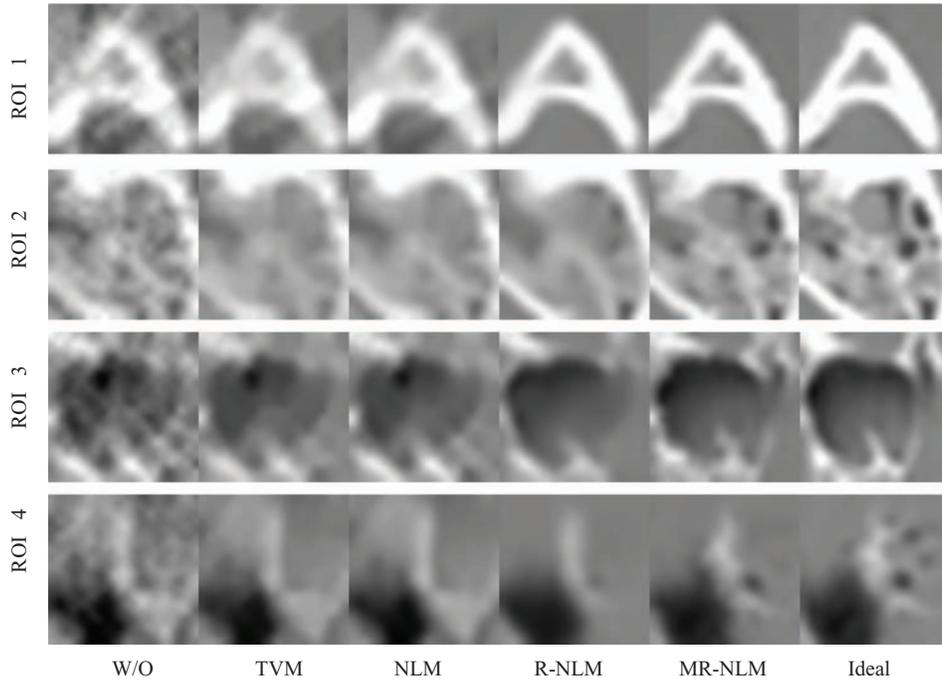


FIG. 11. Comparing the four ROIs defined in Fig. 9 visually.

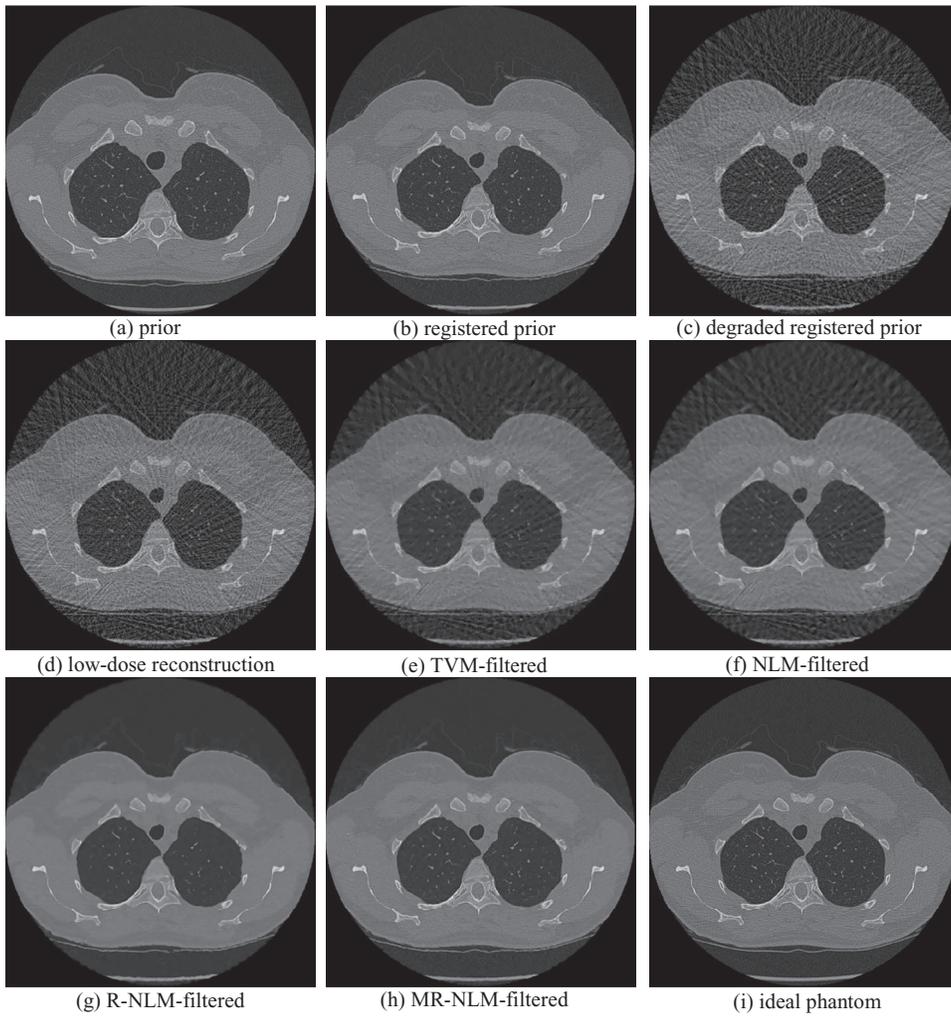


FIG. 12. Results obtained for the human lung.

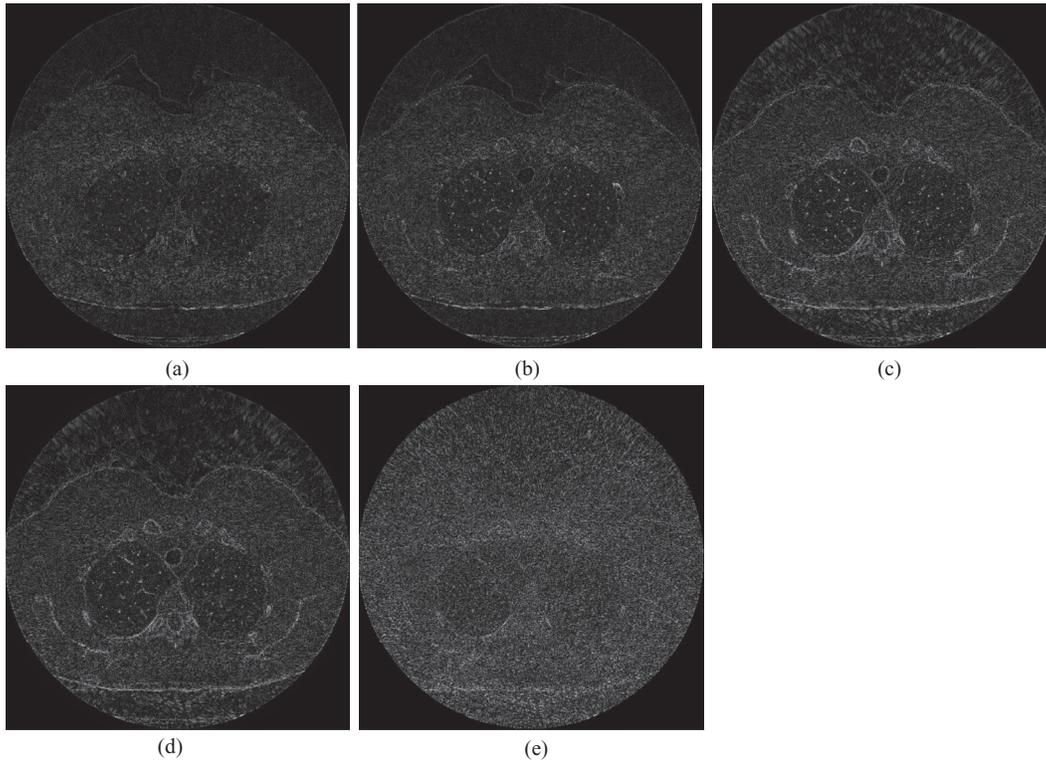


FIG. 13. Absolute difference images for the human lung. (a) MR-NLM; (b) R-NLM; (c) NLM; (d) TVM; and (e) without filtering.

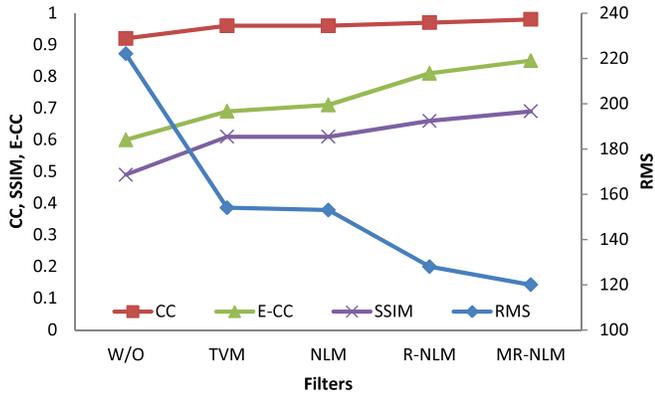


FIG. 14. Graphical comparison of the results obtained for the human lung via various metrics.

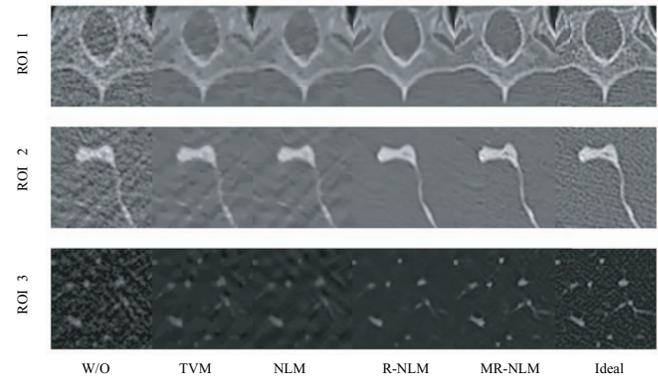
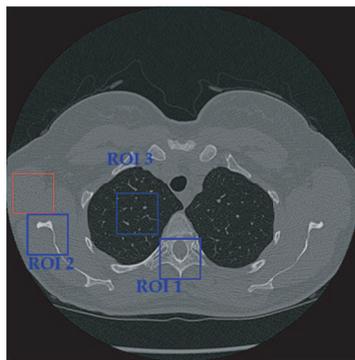
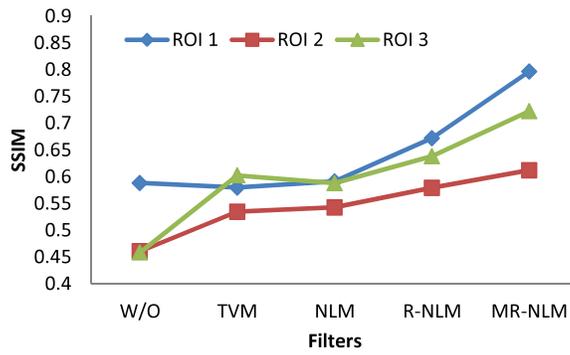


FIG. 16. Comparing the two ROIs defined in Fig. 14 visually.



(a)



(b)

FIG. 15. ROI-based analysis for one slice of the human lung. (a) ROI locations and (b) SSIM evaluation.

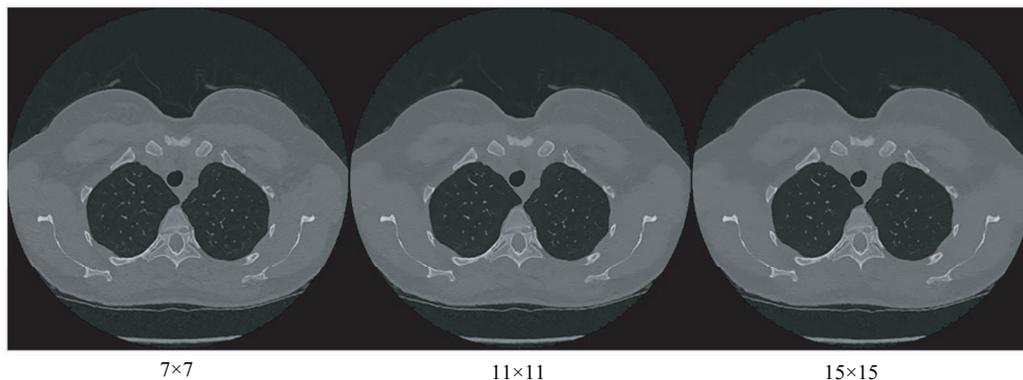


FIG. 17. Comparing the results obtained with three search window sizes.

depends on how well the registration performs and how much the overall structure changes between low-dose and normal-dose scan. A larger size results in more smoothness and a reduction of detail, but it also increases robustness when the registration is not perfect. As mentioned, we have used a 7×7 window for all experiments and as this figure demonstrates this window size provides a good trade-off on smoothness and detail preservation.

IV. CONCLUSIONS

We have demonstrated an efficient noniterative framework for low-dose CT image reconstruction that utilizes an available prior regular-dose scan to assist in the NLM-based regularization of a filtered backprojection reconstruction plagued with significant low-dose artifacts. We have specifically addressed the case when dose reduction is achieved with a lesser number of projections which typically results in severe streak artifacts. Therefore, the reduction of the dose is directly related to the reduction of projections. We find that a crucial element in this effort is to simulate the same low-dose artifacts also in the registered prior to facilitate a more accurate structure matching for subsequent regularization with samples in the registered clean prior. The overall purpose of this work is to make the low-dose image faster readable by reducing the streak artifacts and increase the visibility of the features. While all of these image features can also be seen in the low-dose image, recognizing them requires a time-intensive visual inspection which reduces diagnostic throughput and also makes clinical reasoning much more difficult.

The main limitation in using our method for reducing the number of views is the registration of the prior with the degraded reconstruction image. We have chosen the SIFT-flow method which we found to perform better than the Demon algorithm²⁸ in the presence of noisy data. We suspect that this might be because SIFT-flow uses a structure-sensitive feature descriptor at multiple scales and might ignore noise artifacts better. One item of future work is to test other registration methods and see if they perform even better and so allow a further reduction of projections. It would also be interesting to use the SIFT-flow registration technique for noisy reconstructions obtained with low mA or kV settings and compare

the outcome with those obtained using the EMP-MI approach of Ma *et al.*¹⁴

Another topic of study is to see how sensitive the method is with regards to newly appearing or vastly changing features in the follow-up low-dose scan. Our NLM-based matching scheme is designed to fall back to conventional NLM if no reliable match can be found in the registered neighborhood. Future work will study this fail-safe design using a wide set of structures in a rigorous fashion. Further, we also aim to perform a detailed study with regular-dose and subsequent low-dose projection data directly obtained from a scanner, which we did not have access to for this present work. But nevertheless, we believe that our results clearly demonstrate the conceptual merit of our method.

Further, we find that the perceptual image quality metrics track quite well what can also be visually observed from the reconstructed images. Specifically, we find that our efficiently computed E-CC metric shows similar trends as the more computationally involved SSIM metric. It therefore represents a good alternative when image evaluation must be fast. For future work, we plan to involve clinicians in the image assessment to further validate the suitability of E-CC and SSIM for our purposes.

Current work also focuses on accelerating the R-NLM and MR-NLM frameworks entirely on the GPU in order to increase appeal to clinical applications, also in conjunction with our clinician-based validation and fine tuning.

ACKNOWLEDGMENTS

This work was funded in part by National Science Foundation (NSF) EAGER Grant No. 1050477 and NSF Grant No. 0959979, and by the IT Consilience Creative Program sponsored by the Ministry of Korea Knowledge Economy, Republic of Korea. We also thank the Stony Brook Radiology Department for scanning the phantom dataset and making the scans available to us.

^{a)}Electronic mail: wxu@cs.sunysb.edu

^{b)}Electronic mail: mueller@cs.sunysb.edu

¹D. Brenner and E. Hall, "Computed tomography: An increasing source of radiation exposure," *N. Engl. J. Med.* **357**, 2277–2284 (2007).

- ²F. Mettler, B. Thomadsen, M. Bhargavan, D. Gilley, J. Gray, J. Lipoti, J. McCrohan, T. Yoshizumi, and M. Mahesh, "Medical radiation exposure in the U.S. in 2006: Preliminary results," *Health Phys.* **95**(5), 502–507 (2008).
- ³F. Natterer, *The Mathematics of Computerized Tomography* (John Wiley, New York, 1986).
- ⁴L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, 259–268 (1992).
- ⁵E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory* **52**, 489–509 (2006).
- ⁶E. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.* **53**(17), 4777–4807 (2008).
- ⁷H. Yu and G. Wang, "A soft-threshold filtering approach for reconstruction from a limited number of projections," *Phys. Med. Biol.* **55**(13), 3905–3916 (2010).
- ⁸R. Gordon, R. Bender, and G. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," *J. Theor. Biol.* **29**, 471–481 (1970).
- ⁹A. Andersen and A. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," *Ultrasound. Imaging* **6**, 81–94 (1984).
- ¹⁰P. Gilbert, "Iterative methods for the 3D reconstruction of an object from projections," *J. Theor. Biol.* **76**, 105–117 (1972).
- ¹¹C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(5), 978–994 (2011).
- ¹²Z. Kelm, D. Blezek, B. Bartholmai, and B. Erickson, "Optimizing non-local means for denoising low-dose CT," *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '09)*, pp. 662–665 (2009).
- ¹³H. Yu, S. Zhao, E. Hoffman, and G. Wang, "Ultra-low dose lung CT perfusion regularized by a previous scan," *Acad. Radiol.* **16**, 363–373 (2009).
- ¹⁴J. Ma, J. Huang, Q. Feng, H. Zhang, H. Lu, Z. Liang, and W. Chen, "Low-dose computed tomography image restoration using previous normal-dose scan," *Med. Phys.* **38**(10), 5713–5731 (2011).
- ¹⁵Z. Lu, Q. Feng, P. Shi, and W. Chen, "A fast 3-D medical image registration algorithm based on equivalent meridian plane," in *Proceedings of the IEEE International Conference on Image Processing* (2007), pp. 357–360.
- ¹⁶W. Xu and K. Mueller, "A reference image database approach for NLM filter-regularized CT reconstruction," in *Proceedings of the Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, Potsdam, Germany (July, 2011), pp. 116–119.
- ¹⁷A. Buades, B. Coll, and J. Morel, "A review of image denoising algorithms with a new one," *Multiscale Model. Simul.* **4**(2), 490–530 (2005).
- ¹⁸W. Xu and K. Mueller, "Evaluating popular non-linear image processing filters for their use in regularized iterative CT," in *Proceedings of the IEEE Medical Imaging Conference*, Knoxville, TN, October 2010.
- ¹⁹J. Bian, J. Siewerdsen, X. Han, E. Sidky, J. Prince, C. Pelizzari, and X. Pan, "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT," *Phys. Med. Biol.* **55**, 6575–6599 (2010).
- ²⁰D. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision* (1999), pp. 1150–1157.
- ²¹<http://people.csail.mit.edu/celiu/SIFTflow>.
- ²²Z. Zheng, W. Xu, and K. Mueller, "Performance tuning for CUDA-accelerated neighborhood denoising filters," in *Proceedings of the 3rd Workshop on High Performance Image Reconstruction* (held with Fully 3D 2011), Potsdam, Germany (July, 2011), pp. 52–55.
- ²³F. Xu and K. Mueller, "Real-time 3D computed tomographic reconstruction using commodity graphics hardware," *Phys. Med. Biol.* **52**, 3405–3419 (2007).
- ²⁴W. Xu and K. Mueller, "Learning effective parameter settings for iterative CT reconstruction algorithms," in *Proceedings of the Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, Beijing, China, September 2009.
- ²⁵Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
- ²⁶Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.* **9**(3), 81–84 (2002).
- ²⁷W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Represent* **22**(4), 297–312 (2011).
- ²⁸J. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Image Anal.* **2**, 243–260 (1998).
- ²⁹<http://www.imagegently.org>.
- ³⁰<http://www.imagewisely.org>.
- ³¹<http://www.giveascan.org>.