Database-assisted low-dose CT image restoration

Wei Xu,[a] Sungsoo Ha,[a] and Klaus Mueller[a]
Center of Visual Computing, Computer Science Department, Stony Brook University, Stony Brook, New York 11794-4400

(Received 2 August 2012; revised 14 January 2013; accepted for publication 15 January 2013; published 28 February 2013)

Purpose: Acquiring data for CT at low radiation doses has become a pressing goal. Unfortunately, the reduced data quality adversely affects the quality of the reconstructions, impeding their readability. In previous work, the authors showed how a prior regular-dose scan of the same patient can efficiently be used to mitigate low-dose artifacts. However, since a prior is not always available, the authors now extend the authors' method to use a database of images of other patients.
Methods: The authors' framework first matches the low-dose (target) scan with the images in the database and then selects a set of images that contain anatomical content similar to the target. These "priors" are then registered to the target and form the set of regular-dose priors for restoration via an extended nonlocal means (NLM) filtering framework. To accommodate the larger spatial variability of the patient scans, the authors subdivide the image area into blocks and perform the filtering locally. The database itself is first preprocessed to map each image from its 2D image space to a corresponding high-D image feature space. From this encoding a visual vocabulary is learned that assists in the query of the database.
Results: The authors demonstrate the authors' framework via a lung scan example, for both streak artifacts (resulting from smaller projection sets) as well as noise artifacts (resulting from lower mA settings). The authors find that in the authors' particular example case three priors were sufficient to restore all features faithfully. The authors also observe that the authors' method is quite robust in that it generates good results even when the noise conditions significantly worsen (here by 20%). Finally, the authors find that the restoration quality is significantly better than with conventional NLM filtering.
Conclusions: The authors image restoration algorithm successfully restores images to high quality when the registration is well performed and also when the priors match the target well. When the priors do not contain sufficient information, the affected image regions can only be restored to the quality achieved with conventional regularization. Hence, a sufficiently rich database is a key for successful artifact mitigation with this approach. Finally, the blockwise scheme demonstrates the potential of using small patches of images to form the database. [http://dx.doi.org/10.1118/1.4790693]

Key words: image restoration, artifact mitigation, low-dose CT, nonlocal means, database-assisted, prior-based

I. INTRODUCTION

Motivated by the need to minimize the radiation exposed to patients, a growing number of research efforts have been dedicated to the topic of *low-dose* CT. Lowering the radiation dose can be achieved either by reducing the number of x rays, their energy, or both. However, a direct effect of these dose reduction efforts is CT images with strong noise artifacts, streaks, and reduced feature detail—all of which impede image readability in diagnostic tasks. To overcome these problems one can either apply iterative reconstruction schemes with the goal of optimizing the reconstruction given the limited data[1–3] or one can try to reduce the artifacts in the image domain via a suitable image restoration method.[4]

For the latter option, neighborhood filters, in particular, the nonlocal means (NLM) filter,[5] have shown great promise for the restoration of degraded low-dose CT imagery.[4,6] Originally devised for general image denoising tasks, NLM is essentially an extended Gaussian filter. It updates a given pixel by looking for pixels with statistically similar local neighborhoods in the image and then Gaussian-weighs their contributions by the degree of similarity. The extent of the search is specified by a *search window*, while the size of the neighborhood used for similarity matching is called a *neighborhood patch*. A more recent trend in CT reconstruction has been to extend the search window beyond the image subject to restoration. Schemes have been devised that utilize a prior scan[7–9] of the same patient to search for high-quality updates. Other work has successfully constrained a reconstruction by images of the same dynamic scan.[1] While all this produces excellent results, such a prior scan or dynamic scan of the same patient may not always be available.

To meet this inherent shortcoming we propose the idea of extending the search window even further, namely, to a collection of images of *different* patients. This approach, in fact, is quite alike the psychophysical processes that occur in medical professionals when viewing degraded imagery. They also borrow from their extensive medical training and experiences

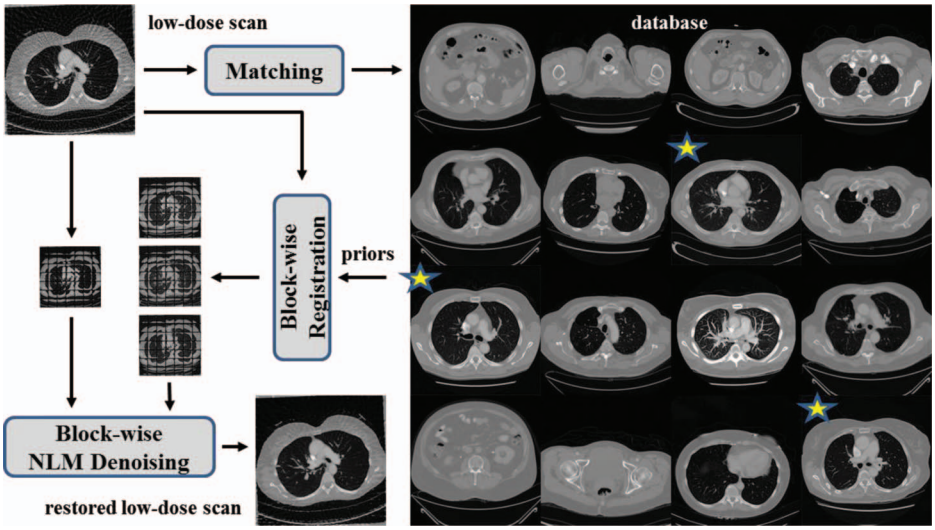

FIG. 1. Illustration of our framework by ways of a (small) lung database example.

to see the "true patterns behind the noise." These recognition tasks, however, take valuable time and can also lead to frustration, and it is for this and other reasons that artifact reduction by image processing and algorithmic means is an important mission.

Using collections of clean images to reduce noise or blur in degraded images is not a new idea, at least not in general. There are in fact two rather disjoint schools of thought, and both aim to cope with the extremely large space of possible image detail. The first approach first constructs a large-scale database of possible detail at some level and then uses a sophisticated matching strategy to retrieve the detail of interest from this database.[10] The other approach is based on *sparse coding*. It first constructs a dictionary of representative base patterns which must then be optimally combined for reproducing the desired detail of interest.[11] While the first approach is a top-down search, the second is bottom-up. Both strategies can be justified by theories on how humans perform visual search, which likely is a conjunction of both.[12]

Our paper expands on two workshop papers[13,14] where we have presented preliminary thoughts as well as encouraging experimental results using the first—the database—approach for low-dose artifact mitigation. Parallel to our work, another team of researchers[15,16] has pursued with similar success the alternative—the sparse coding—approach for the same purpose. Since the vote is still out, on an even grander scale, about which of the two strategies is better or more likely, we refrain from making such claims here. Our sole purpose is rather to formalize the framework we conceived, expose results on what is currently possible when using it for low-dose CT, and point out current shortcomings that warrant future work.

Figure 1 illustrates our framework by ways of an example: the restoration of a low-dose lung (target) scan using a database of regular-dose lung images. First, we match the target scan with the images in the database and select a set of images (marked with stars) containing similar anatomical content as the target. We then register these images to the target to form the set of artifact-free priors. Finally, using these priors we apply the extended NLM-filter scheme to denoise the target via a blockwise update strategy.

Our paper is organized as follows. Section II describes both methodology and technical detail, Sec. III presents results, and Sec. IV ends with conclusions and future work.

II. METHODS AND MATERIALS

The workflow of our method consists of three major stages: offline database construction, online prior-search, and online denoising (the latter two stages are illustrated in Fig. 2). In the offline database construction stage, we create a global image feature descriptor G to represent each image in the given image database. This forms the global feature database. A visual vocabulary V summarizing the local image features is also learned in this stage. Then in the online prior-search, we generate $G(I)$ with V for the target image I. Following, we use $G(I)$ to query the global feature database to find the M nearest neighbors as regular-dose priors. These priors have (artifact-free) anatomical content that is most similar to the (degraded) anatomical content of the target. Next, in the online denoising, we first align the regular-dose priors to the target in a blockwise manner. These images form the set of clean registered prior (CRP) blocks. Using these blocks, we run what we call Reference-based NLM (R-NLM),[8] where we use the prior blocks for NLM matching and look up the pixel values in the corresponding CRP block. We now describe each of these components in closer detail.

II.A. Local image feature descriptor

Image matching is a fundamental operation in computer vision and image processing and it is often used for scene and object recognition. Typically, the image is expressed as a high-dimensional feature vector and the matching occurs in this high-D feature space. Since we wish to match image features in the presence of significant noise and streaks

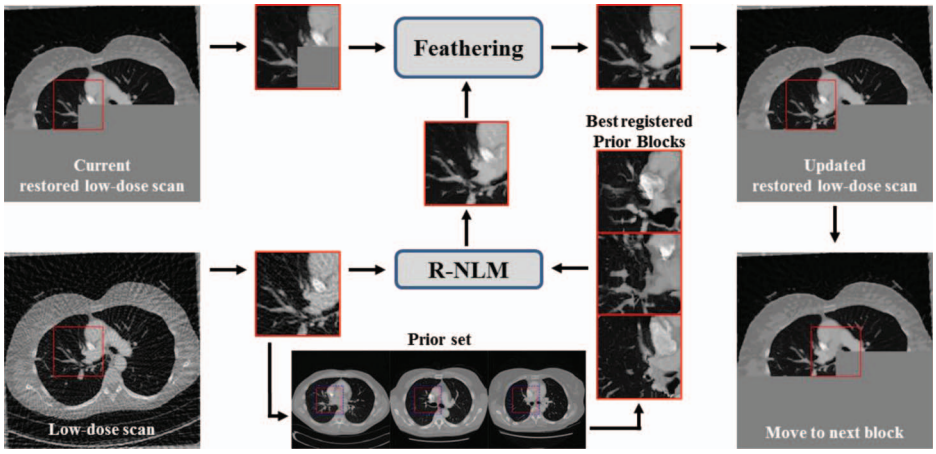

FIG. 2. Blockwise registration and R-NLM denoising workflow.

artifacts, we require a feature descriptor that is insensitive to these degradations. The scale-invariant feature transform (SIFT) (Ref. 17) is such a feature descriptor. It captures the histogram of edges in a local neighborhood at multiple levels of scale, characterizes salient local and transform-invariant image structures, and encodes contextual information. A SIFT feature descriptor is usually a 128-D vector encoding 8-orientation histograms of edges over 4×4 blocks with each block of size 4×4, serving as a local descriptor of the image. In its original definition, only keypoint locations are selected. However, it was shown that dense SIFT vectors (dSIFT) on a regular spaced grid could provide more correspondence and reveal more details also in flat texture areas and are thus more robust.[18,19]

We found that a grid spacing of 8 pixels works well for dense SIFT in our application. Thus, for an image of size 256^2 we get 32×32 SIFT vectors, while for size 512^2 we get 64×64 SIFT vectors. We use these local image features to compute the visual vocabulary, as shown next.

II.B. Global image feature formation

To form a global image feature descriptor from local ones, traditional dense SIFT algorithms follow the bag-of-feature method.[20] We use the following algorithm to accomplish this:

(1) Extract the local feature descriptors: generate a set of SIFT local feature descriptors $\{S_0, S_1, \ldots, S_{N-1}\}$ to represent each image.
(2) Build the visual vocabulary: randomly select the local feature descriptors of all images in the database and perform k-means clustering to learn K cluster centers as visual words $\{V_0, V_1, \ldots, V_{K-1}\}$. This forms the visual vocabulary V of the database.
(3) Label the local features to the visual words: for each image, assign the index of the closest visual word to each local feature vector.
(4) Perform vector quantization to generate a global feature descriptor: compute the histogram of visual words in each image $\{H_0, H_1, \ldots, H_{K-1}\}$ and concatenate the weighted histogram series into a long vector to form the global feature descriptor.

One drawback of this method is that the feature's location information in the original 2D image space is discarded. To make use of this spatial information and keep track of it in multiresolution, we exploit a spatial pyramid scheme[18] to implement a "stronger" feature description. The multiresolution layers are formed by recursively subdividing the image into $b \times b$ blocks. In a layer L, for each block, only the feature vector extracted from that block is aggregated into the histogram of its specific visual word. In this way, the clustering is still performed in feature space while the histogram pyramid is built in 2D image space. The weight for each histogram is inversely proportional to its block width. We call the resulting histogram sets the *spatial pyramid-based histograms*.

There are four parameters associated with this part: the number of visual words K, the number of layers L, and the block size b. We used 60 visual words of the local feature descriptors, and each image was represented with a spatial pyramid-based histogram composed of one layer and 5×5 blocks. These settings showed the best query precision in the online prior search stage. Sec. II.C provides further detail.

II.C. Online prior search

In this stage, the global feature vector of the target scan is generated online following the same steps as outlined for the offline stage. This long feature vector is then used to search for similar priors in the database. These priors anatomically characterize the same content as the target scan but may contain small variations in scale, rotation, and deformation. We found that histogram intersection performs better than the Euclidean distance for gauging the similarity for matching. The histogram intersection operates within a spatial pyramid, i.e., the intersection is counted both blockwise and visual word-wise and is then summed up to form a single value.[18] To ensure that the priors contain a wide variety of diverse anatomical features, we perform the searching process patient by patient. We then construct a ranked-list by selecting and sorting only the top ranked priors in each patient. Finally, additional

search within the top-ranked priors in the list further expands and refines the list.[21]

II.D. Registration

Once the regular-dose prior (or reference) scans are found, the online denoising process can be executed. The first step is to register the priors with the target scan. Different from our initial work[13] we perform both the registration and the denoising in a blockwise fashion. This provides for better local control which is needed since the database priors have less correspondence to the target than priors coming from the same patient. More specifically, we create a small block of size 129 × 129 and shift that block with a step size of 64 in raster-scan order. Figure 2 illustrates this process. For each block, we perform the registration by aligning that block with the corresponding blocks in the prior scans (red boxes in Fig. 2). This local registration relaxes the strict requirements of a global image registration and allows for priors to only partially match the target, which is likely since they come from different patients.

We used the SIFT-flow registration algorithm[19] for the registration. SIFT-flow is a state-of-the-art registration method that originates from the optical-flow algorithm and produces dense, pixel-to-pixel correspondences between two images. It extends optical-flow matching from raw pixels to SIFT features, which significantly improves robustness when registering artifact-rich with artifact-free images. In order to further improve the quality of the registration we predenoise the target scan via Gaussian filtering.

II.E. R-NLM denoising

R-NLM (Ref. 13) follows the standard NLM filtering scheme but uses the artifact-free registered prior images, CRP, instead of the target itself. Thus, the pixel weights are computed by comparing patches in the target with patches in the prior images. More formally,

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

Here, x is the location of the target pixel and 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 L_2 distance between two patch vectors with t representing the index within a patch. The factor h controls the overall smoothness of the filtering. In our case h is larger than typically used for standard NLM to accommodate higher noise level. The superscript CRP indicates that the pixels originate from the artifact-free registered priors CRP.

Equation (1) differs from the one used in Ref. 8 in that we removed the Gaussian weighting in the patch similarity measurement. We found in experiments that this leads to better matches since a greater patch neighborhood influences it. The direct consequence of better matches is an increased sharpness at edges, as demonstrated in Ref. 8 for the matched artifact MR-NLM scheme. We could not use MR-NLM in the work presented here since it proofed difficult to simulate realistic streak artifacts in a block. Noise would have been easier but we strived for a scheme that applies to both types of artifacts unilaterally. Eliminating the Gaussian kernel and replacing it by a wide box filter seems to better "see through" the noise and capture the true pattern underneath more faithfully.

For pixels for which no similar patch can be found within the search window we perform standard NLM with a smaller h. The pixels are detected by comparing their denominators in Eq. (1), which represents the summation of contributions from pixels in the search window, with a preset threshold—we used 0.3. For these pixels the denoising falls back to standard NLM. This conservative approach ensures that no false ill-fitting features are introduced.

We perform the R-NLM in a blockwise, raster-scan fashion, as shown in Fig. 2. For each block, when the R-NLM has finished, the pixels in the overlapping regions are *feathered* in the raster-scan order. Feathering retains the edges but adjusts the grey levels such that newly added regions blend well with existing ones.[22] It allows one to not only retain local contrast among neighbor but also to remove remaining noise at some loss of sharpness.

III. EXPERIMENTAL RESULTS

We constructed a human lung database of 30 patients (41 138 512^2 images) from an online human lung database (http://www.giveascan.org). The images were not prealigned. To match the quality of the CT scans, all scans were regenerated from 720 projections over 360° with a fan-beam geometry (fan angle = 20°). The 720 projections were the point at which there was no more quality improvement, as gauged by RMSE with the original scan. We used these scans as the gold standard in all experiments. We then picked subsets of these projections to generate the low-dose scans with streak artifacts. To simulate the noise artifacts, we added various levels of noise [dB, signal-to-noise ratio (SNR)] to the sinogram of the gold standard images. To create a new scan different from any scan in the database (even if they were already from different patients), the selected scan was first deformed or rotated (to mimic a real clinical situation), forward projected, and then reconstructed with the low-dose condition under study.

In Secs. III.A and III.B, we report on two experiments to evaluate and validate the proposed algorithm. In all experiments, we used a patch size of 7 × 7 with a 13 × 13 search window for the NLM filter. Then, both the overall smoothness parameter, h, and the threshold parameter which decides between standard NLM and R-NLM were experimentally chosen by inspecting the quality of the restored image for each streak and noise reduction task.

III.A. Priors quality and diversity

We performed experiments at which the database-assisted restoration occurred at three levels of difficulty, gauged as a function of closeness of the available priors to the target anatomy. In these experiments the database contained, among

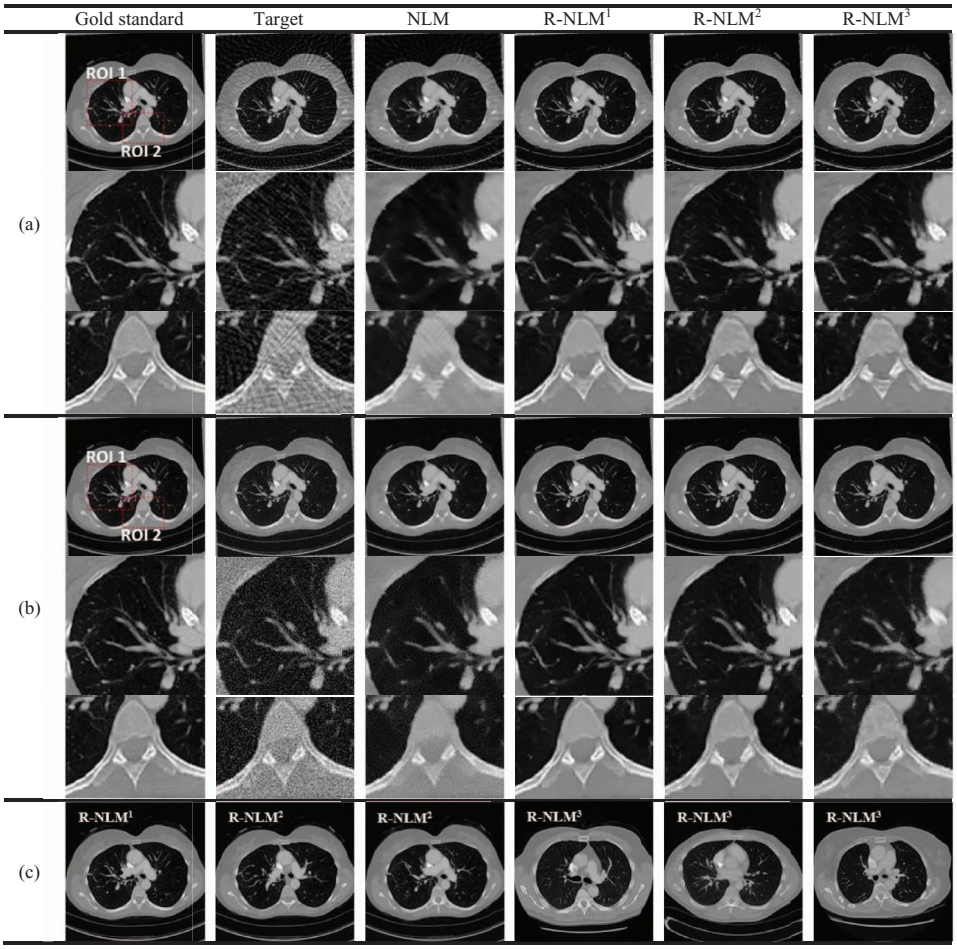

FIG. 3. Restoration results for (a) streaks (86 projections) and (b) noise (30 SNR dB) using (c) different sets of priors. R-NLM^1 is a prior image from the same patient—this image prior is from scan of the same slice and thus very similar to the target. R-NLM^2 are two different image priors also from the same patient but from different slices and so less similar to the target. R-NLM^3 are three priors from three different patients located automatically in our database. The corresponding restorations are shown in rows (a) and (b) along with results obtained with the standard NLM approach.

other scans: (1) one almost identical CT scan of the same patient; (2) two somewhat similar CT scans of the same patient; and (3) only CT scans of other patients. Target images were generated at the following two low dose conditions: (1) reduced data (only 86 projection—this is about 11% of the gold standard data and represents a dose reduction of 88%) and (2) low mA imaging (30 dB SNR Gaussian noise was added to the sinogram). Figure 3 shows the restoration results for these two conditions using the different types of priors described above, and compares them with results obtained with the standard NLM method. We observe that the prior-based scheme significantly improves image quality even with "foreign"

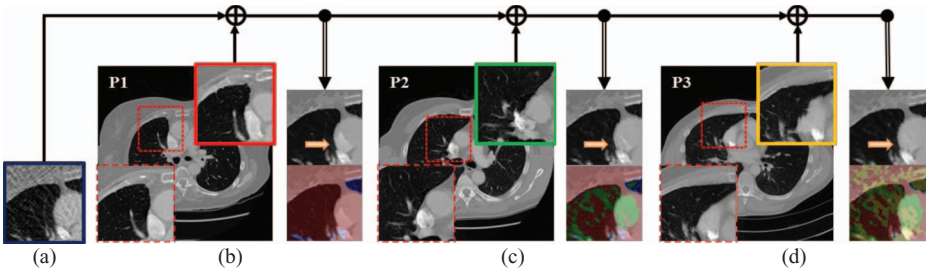

FIG. 4. Multiple priors: effect and benefits. By increasing the number of priors used for scan restoration (here, streak reduction), we gain a much wider range of anatomical features. This prevents the borrowing of pixel values from wrong structures or the failure of finding structures at all, during the restoration process. (a) Target block subject to denoising. [(b)–(d)] Restoration using: (b) only prior P1, (c) prior P1 and P2, and (d) priors P1, P2, and P3. In each subfigure, the prior is shown with the matched block marked by a red-dotted outline. The two inserted (zoomed) images are: (bottom left) the original block and (top right) the block aligned to (a). The two stacked images to the immediate right are (top) the restoration result and (bottom) the same block with the pixels colored by the prior from which they originate. The (blue, green, yellow) pixels come from prior (P1, P2, P3). We observe that only when using all three priors the structure pointed to by the arrow gets restored in the most plausible way, according to its noisy counterpart in (a).

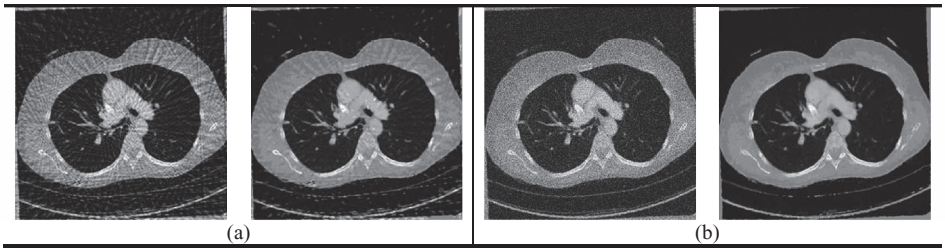

FIG. 5. Robustness of (a) streak and (b) noise reduction: The restorations use the three priors introduced in Fig. 3 (denoted R-NLM^3). The target images are generated with 20% less projections (70) (a, left) or 20% more noise (25 SNR dB) (b, left) than in the study of Fig. 3. The restored images for each (a) and (b) are shown on the target's right.

priors from different patients. The edges are sharper and detail is better preserved.

Figure 4 shows that having a more diverse set of priors can improve the outcome tremendously. In the example given, we needed at least three priors to successfully restore the structure pointed to by the arrow. This is not unlike the case in which a more experienced radiologist "reads" a noisy image. And indeed, the need for a massive database has been confirmed in research that aimed to remove unwanted structures in photographs.[10] Constructing such a large database is our current goal, with a need for "big data" management.

III.B. Data quality

To test robustness, we lowered the quality of the data by approximately 20% and observed the restoration outcome obtained with a database that only contained scans from other patients (case 3 above). In one experiment we reduced the number of projections from 86 to 70, which is a further dose reduction of about 20%. In the other experiment we increased the level of sinogram noise further, from 30 to 25 dB SNR. To calculate the decrease in dose we can use the relationship $\text{SNR} = N/\sigma = N/\sqrt{N} = \sqrt{N}$, where σ is the level of quantum noise and N is the number of x-ray photons. Since the dose is directly and linearly related to N we achieve a further reduction in dose by 30%. Our experimental results are shown in Fig. 5. We observe that the outcome is still acceptable—all major features are still well preserved. This is especially true for the noise case, while the streak case would probably benefit from an alternate iterative CT reconstruction scheme such as SART (as opposed to the current FDK).

IV. CONCLUSIONS

We proposed a general framework for high quality restoration of low-dose CT scans with the help of a general CT image database. We believe our approach is attractive because once the database has been established the online restoration process is quite fast. The restoration process itself, once the priors have been selected from the database, is just a minor extension to standard NLM filtering which is easy to implement and efficient to run. Additional efficiencies have resulted from performing the reconstruction and the restoration steps sequentially. While this is not uncommon (see, e.g., Refs. 2, 7, and 8), there is a chance that separating these two processes results in artifacts and loss of detail. In order to study these effects further we plan to compare our approach with the one by Xu *et al.*[15] who combine these two steps into one.

Our results point out that a sufficiently elaborate database is crucial to the success of our method. Since modern PACS systems now have massive CT data on cheap disks we do not see this as a major obstacle. Future work will focus on enriching our database with more data, also of pathologies, and so create a system that requires our conservative fallback NLM scheme only in rare cases. In addition, in order to still maintain a manageable set of priors we are currently working on transforming our present image-based prior set to a patch-based set. This will eliminate redundancies and provide for a more diverse feature set to base the matching on. Also, we intentionally did not embed our method into an iterative reconstruction pipeline such as SART or ASD-POCS.[3] We wanted to test how far a single restoration step can take us. Next, we will iterate the databases-assisted restoration with a data-driven reconstruction which will likely improve the results further and also serve for verification.

We close by stating that the foremost purpose of a CT scan is to gain insight into a patient's state of health. Low-dose imaging is also important, but it should not be at the expense of increasing the likelihood of false positives or negatives. Our approach uses techniques from machine learning to improve the quality of low-dose CT scans. It grows by the quality of the examples it is taught with and the sophistication of its algorithms. Without sounding too futuristic, this is not too dissimilar from the educational process of a radiologist. We have clearly a long way to go, and we will likely never be able to match the tremendous capabilities of the human brain, but we may achieve a reliable digital doctor's assistant.

ACKNOWLEDGMENTS

This work was funded in part by National Science Foundation (NSF) Grant Nos. IIS-1050477, CNS-0959979, and IIS-1117132. The authors also thank the Stony Brook Radiology Department for making CT scans available to us.

[a]Authors to whom correspondence should be addressed. Electronic addresses: wxu@cs.sunysb.edu; sunha@cs.sunysb.edu; and mueller@cs.sunysb.edu

[1]G. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," Med. Phys. **35**(2), 660–663 (2008).

[2]X. Jia, Y. Lou, R. Li, W. Song, and S. Jiang, "GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation," Med. Phys. **37**, 3441–3447 (2010).
[3]E. 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).
[4]Z. Li, L. Yu, J. Trzasko, J. Fletcher, C. McCollough, and A. Manduca, "Adaptive nonlocal means filtering based on local noise level for CT denoising," Proc. SPIE Medical Imaging, 8313H (2012).
[5]A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," Multiscale Model. Simul. **4**(2), 490–530 (2005).
[6]W. Xu and K. Mueller, "Evaluating popular non-linear image processing filters for their use in regularized iterative CT," in *IEEE Medical Imaging Conference*, October 2010.
[7]J. 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), 5713–5732 (2011).
[8]W. Xu and K. Mueller, "Efficient low-dose CT artifact mitigation using an artifact-matched prior scan," Med. Phys. **39**(8), 4748–4760 (2012).
[9]H. Yu, S. Zhao, E. Hoffman, and G. Wang, "Ultra-low dose lung CT perfusion regularized by a previous scan," Acad. Radiol. **16**(3), 363–373 (2009).
[10]J. Hays and A. Efros, "Scene completion using millions of photographs," ACM Trans. Graphics **26**(3), 4 (2007).
[11]M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Process. **15**(12), 3736–3745 (2006).
[12]L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998).
[13]W. Xu and K. Mueller, "A reference image database approach for NLM filter-regularized CT reconstruction," in *Proceedings of the Fully3D*, 2011, pp. 116–119.
[14]W. Xu, S. Ha, and K. Mueller, "Database-assisted low-dose CT image restoration," in *Proceedings of the 2nd International Conference on Image Formation in X-ray Computed Tomography*, Salt Lake City, UT, July 2012, pp. 111–114.
[15]Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Dictionary learning based low-dose x-ray CT reconstruction," *Proceedings of Fully 3D*, Germany, 2011, pp. 258–261.
[16]L. Yang, J. Zhao, and G. Wang, "Few-view image reconstruction with dual dictionaries," Phys. Med. Biol. **57**(1), 173–189 (2012).
[17]D. Lowe, "Object recognition from local scale-invariant features," Proc. IEEE Int'l Conf. Comp. Vision **2**, 1150–1157 (1999).
[18]S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the Computer Vision and Pattern Recognition* (2006), Vol. 2, pp. 2169–2178.
[19]C. 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).
[20]G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proceedings of the ECCV Workshop on Statistical Learning in Computer Vision*, 2004.
[21]O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," in *Proceedings of the International Conference on Computer Vision*, 2007.
[22]P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," ACM Trans. Graphics **22**(3), 313–318 (2003).