

Database-Assisted Low-Dose CT Image Restoration

Wei Xu, Sungsoo Ha and Klaus Mueller, *Senior Member, IEEE*

Abstract– The image quality of low-dose CT scans typically suffers greatly from the limited utilization of X-ray radiation. Although the harmful effects to patient health are reduced, the low quality of the reconstructions makes diagnostics difficult. In previous work, we have demonstrated a method that can restore a low-dose image by ways of a database of reference images. This database stored a set of pre-aligned non- and pre-corrupted reference CT images to support a matched-reference non-local means (MR-NLM) filtering approach. While effective, the need to store images with many different types of corruptions and alignments greatly impeded system scalability. In this current work, we have significantly simplified the database which now is comprised of just a set of regular-dose patient scans. Our present scheme performs both alignment and artifact generation on the fly and uses a sophisticated image and feature matching scheme to find good candidates to support our MR-NLM filtering scheme.

I. INTRODUCTION

In recent years a growing amount of research has been dedicated to low-dose CT, motivated by the need to minimize the radiation exposed to patients while maximizing the clarity of the reconstructed images to facilitate accurate diagnoses. The adverse low-dose conditions greatly challenge conventional CT reconstruction algorithms, both analytical and iterative. They usually result in images with severe noise artifacts and reduced feature detail. To solve this conundrum, one type of approach enforces better image quality directly in the reconstruction process [6][12][17], while another improves the image quality in a post-processing de-noising step [7]. Our paper belongs to the second category.

Neighborhood filters, in particular non-local means (NLM) [1] have shown great promise for the restoration of noisy low-dose CT imagery [17]. To filter a pixel p_i with NLM, its updated value is determined by the values of pixels p_j inside a local neighborhood around p_i , called *search window*. Here, the contribution of a p_j to p_i depends on the similarity between small regions around them, called *patches*. Recently, to extend the search space beyond the current image, some medical imaging researchers have devised schemes that utilize prior scans of the same patient to search for high-quality updates [7][15][16]. We extended these ideas in [18], proposing an approach that utilized an image database of *different* patients which eliminated the need for a prior patient scan. The scheme achieved good artifact mitigation for low-dose scans acquired from only 45 noise-free projections or 60 noisy projections with SNR=10. The database itself contained pairs of artifact-free and artifact-matched reference images. We found that much better filtering results could be achieved by using the latter to find good NLM matches for a contaminated target pixel, but then replacing the noisy target pixel by the

corresponding value from the artifact-free counterpart. We therefore called this method *Matched Reference-Based Non-Local Means* (MR-NLM). A shortcoming of this implementation was that the database could grow rather large since all images needed to be pre-aligned to the target image and also had to contain all possible types of artifacts for matching. In the current paper, we have aimed to reduce these problems and establish a more general framework.

Our present framework embodies a database of regular-dose patient CT scans with no pre-alignment and prior artifact simulation needed. Such scans are commonly available in clinical practice. For image restoration the only assumption we make is that the low-dose CT condition is known. This is reasonable since CT scans are typically obtained following a known reconstruction routine under some geometry configuration with a specific number of projections and mA/kV setting. In the current work we use fan-beam filtered backprojection (FBP) with a limited number of projections with Gaussian noise to simulate the low-dose conditions, but in practice any reconstruction setting can be supported. Our new method still applies the effective MR-NLM scheme, only now we perform alignment and artifact generation on the fly, assisted by a much more sophisticated image and feature matching scheme. We therefore call our framework simply *Database-Assisted CT Image Restoration* (DA-CTIR).

The overall workflow of our method is illustrated in Fig. 1. It consists of three major components:

- **Offline database construction:** given an image database, we create the global image feature descriptor G for each image and build up the global feature database. A visual vocabulary V is also learned.
- **Online prior search:** for the input image I , generate $G(I)$ with V and use it to query the global feature database to find the M nearest neighbors (NN) as regular-dose *priors*. The priors have the most similar artifact-free content to I .
- **Online de-noising:** align the regular-dose priors to the input image as registered priors (CRP) and corrupt them with the low-dose condition (DRP) to form the prior pairs $\langle CRP_i, DRP_i \rangle$. Finally a refined MR-NLM is performed.

The organization of the paper is as follows. In section II, we describe the methodology including all technical details. Experimental results are presented in section III, followed by conclusions and future work in section IV.

II. METHODOLOGY

A. Local Image Feature Descriptor

Image matching is a fundamental operation in computer vision and image processing and is often used for scene matching and object recognition. An image is usually represented as a high dimensional vector to describe the distinct salient properties of the image. In other words, an image feature

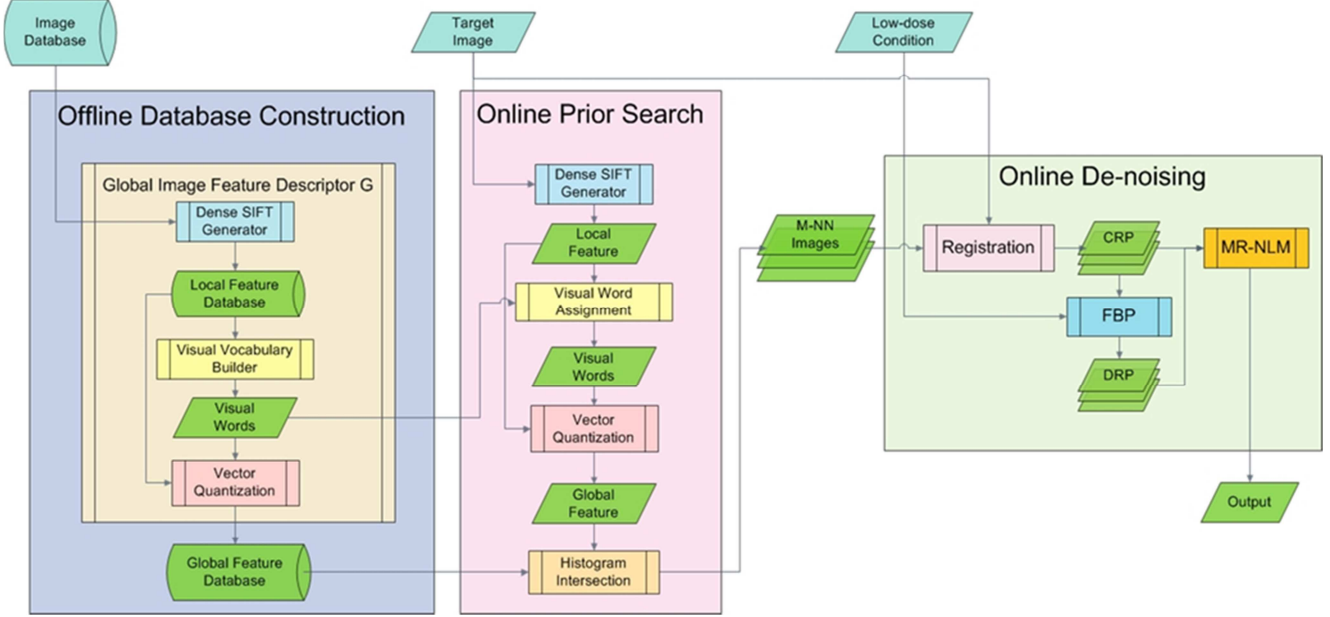


Figure 1: Workflow of the framework: offline database construction, online prior search and online de-noising.

descriptor is employed to map one image from 2D image space to high-D image feature space where image matching is performed. For instance, the GIST descriptor [11] is an aggregated multi-scale oriented edge histogram of the image in a coarse spatial resolution. The Haralick texture features [5] describe the global image statistics based on co-occurrence matrices with different pixel distance values. Although these methods have been shown to work effectively in many applications, they are sensitive to image rotation, distortion and appearance of noise which usually occur in our case.

The scale-invariant feature transform (SIFT) feature descriptor [9], on the contrary, solves these concerns. 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 on a regular spaced grid work better and are more robust [8][10]. Here we also exploit this dense feature scheme so that each image is represented by a fixed number of SIFT vectors.

In this work, we chose a grid spacing of 8 pixels. So for image size 256^2 , 32×32 SIFT vectors are generated while for image size 512^2 there are 64×64 SIFT vectors.

B. Spatial Pyramid Based Global Image Feature Descriptor

To form a global image feature descriptor, traditional dense SIFT algorithms follow the bag-of-feature method [3]. It includes the following steps to combine the local feature vectors into a single one:

- **Extract the local feature descriptors.** Generate a set of SIFT local feature descriptors $\{S_0, S_1, \dots, S_{N-1}\}$ to represent each image.

- **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 $\{V_0, V_1, \dots, V_{K-1}\}$ and so form the visual vocabulary V of the database.
- **Label the local feature descriptors to the visual words.** For each image, its local feature descriptors are assigned to their corresponding closest visual words.
- **Perform vector quantization to generate a global feature descriptor.** Quantize each image's visual words to form histogram series $\{H_0, H_1, \dots, H_{K-1}\}$ of that image. By concatenating the weighted histogram series, a global descriptor is formed.

One drawback of this method is that the feature's location information in the original 2D image space is discarded. To make use of the spatial information and keep track of it in multi-resolution, we exploit a spatial pyramid scheme [8] to implement a "stronger" feature description. The multi-resolution layers are formed by recursively subdividing the image space into $a \times a$ blocks. In a layer L , for each block, only the feature vector extracted from that block is aggregated to the histogram of its specific visual word. In this way, the clustering is still performed in the feature space while the histogram pyramid is built in 2D image space. The weight to each histogram is inversely proportional to its block width.

In this work, for clustering we tried several k values and empirically chose $k=50$ for all databases. This number is relatively small compared to other papers (where $k=200$) which is due to the fact that CT scans are not as complicated as natural images. We set $L=1$ (two layers) and $a=5$ to prevent the splitting of significant body structures [4]. Therefore, for image size 256^2 its global vector dimension is 1,300 while for image size 512^2 it is 2,600.

C. Histogram Intersection and Multiple *kd*-Trees Based Vector Matching

In the *online prior search*, given the learned visual vocabulary V and the computed local features of the target scan, the task of the *visual word assignment* is to find the nearest visual word V_j for each local feature S_i in SIFT vector space (128-D). When processing a set of query images with a large number of dense SIFT vectors, this process could be time consuming. To speed up, we exploit the commonly used *kd*-tree as the nearest neighbor searching data structure. A *kd*-tree is a binary tree that recursively partitions and stores the nodes in k -dimensional space. Counting the number of visited tree nodes is used to measure the complexity of querying the tree.

To handle the query for high-D nodes such as a SIFT vector and reduce backtracking, we employ a multiple principal component *kd*-trees method denoted PKD-trees to perform fast approximate search [13]. In essence, data is first projected onto a PCA-reduced sub-space and arbitrary rotations to data are applied to create multiple trees with different structures. The search order among trees is organized by multithreading. The maximum number of visited nodes is pre-set. We use a Householder matrix as the transformation matrix to speedup arbitrary rotations, and 6 trees are built to accommodate data reduced to 30 dimensions for SIFT vectors.

In the same part, after generating a global feature vector for the target scan, vector matching is performed to search for similar priors in the database that anatomically characterize the same content as the target scan but may contain scale, rotation, and deformation variance. We found that histogram intersection performs better than a Euclidean distance measure. Therefore we implemented the matching with spatial pyramid based histogram intersection which is counted block-wise and visual word-wise and summed up to form a single value [8].

D. Online Denoising

Once the regular-dose prior (or reference) scans have been found, the online de-noising process can be executed. We first register the prior scans with the target scan using the SIFT-flow registration algorithm [10] to make sure the neighborhoods of any pixel position are roughly aligned. Then we reconstruct the artifact matched prior images using the same low-dose condition.

The MR-NLM follows the standard NLM filtering scheme but using a pair of artifact-free and artifact-matched registered prior images $\langle CRP, DRP \rangle$ [18]. More specifically, the weight generation is conducted by comparing patches from the target image and artifact-matched prior images, while the pixel value summation is performed in the corresponding locations in the artifact-free reference images using the weights. The equation of MR-NLM is as follows:

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

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 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 . h controls the overall smoothness of the filtering. The superscript *crp* indicates that the pixels originate from the artifact-free registered prior *CRP*, while *drp* denotes the degraded artifact matched registered prior *DRP*.

In order to further improve de-noising accuracy and enable more efficient computations, we use three refinement strategies. The first two are redundancy control methods designed originally for traditional NLM: (1) reduce patch redundancy by applying PCA to high-D patch space and project patches to a lower dimensional sub-space accordingly [14], (2) reduce search window redundancy by discarding unrelated pixels whose mean and variance values are different enough from the central pixel of the search window in the target image [2], and (3) consider multiple pairs (3 in the experiments) of reference images to broaden the search range.

In this paper, both search and patch windows are of size 7×7 . For the Gaussian kernel, its standard deviation $a = 1$ and the smoothing parameter h is chosen to bring best results.

III. RESULTS

We constructed two databases: a head database (48 256^2 images) by mixing the NIH Visible Human Head (15 images) with a CT cadaver head (33 images) and a human lung database of two patients (150 512^2 images). The images were not pre-aligned. Their original reconstructions were utilized in three different ways. (1) 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 with Gaussian random noise propagated and reconstructed them under the current low-dose condition. (2) We used them to generate an experimental target scan subject to restoration. To create a new scan different from any image in the database, the selected scan was first deformed or rotated (to mimic a real clinical situation), projected, and then reconstructed with the studied low-dose condition. (3) We used the deformed uncorrupted scan to represent the gold standard for evaluation.

A. Performance of the Global Image Feature Descriptor

This experiment was conducted to test the performance of the global image feature descriptor under low-dose conditions. In Fig. 2, both a head scan and a human lung scan were simulated (neither was in the database). I(b) was created by reconstructing a CT head scan after a swirl-like deformation (see I(a)) with 45 projections of SNR 15, and II(b) was created by reconstructing a human lung scan after rotation 5° ccw (see II(a)) with 60 projections of SNR 20. Ideally, the adjacent slices in the same dataset should be found as reference images. The three matched prior images for the head scan are shown as Fig. 2 (c), (d) and (e) and are consistent with our expectations. The case for the lung scan is similar. It confirms that moderate deformations and low-dose artifacts (both streak and noise) do not affect the global feature descriptor to express the underlying anatomical content of the CT images.

B. Performance of PKD-trees Data Structure

For the visual words learned in the first experiment, 6-PKD-tree data structures were created for matching dense SIFT

vectors to the visual words. Two configurations for reducing the dimension of the vectors were set: full 128 and 30. By generating a head scan with various changes such as rotation, resizing, Gaussian noise and affine transform, we extracted 100,000 SIFT vectors from it. As an approximated search, the error rate versus the maximum number of visited nodes M was tested for both dimension settings and is plotted in Fig. 3. When M is above 200, the error rate is lower than 10%. We also observe that dimension reduction of the data does not affect the querying accuracy when the vector is sparse.

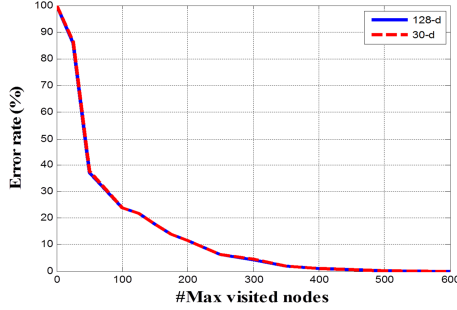


Figure 3: 6-PKD-tree error rate.

C. Performance of Refined MR-NLM

We tested the de-noising effect for both a head image and a human lung scan. In Fig. 2, I(f) and I(g) are the de-noised head image without and with refinement. The lung results are shown as II(c) and II(d) for without and with refinement respectively. For both cases, the basic method restored fine details well. The refined result keeps the same (sometimes better as the area labeled in the box) quality level but reduces the computational complexity greatly.

IV. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed a general framework for high quality restoration of low-dose CT scans with a general CT image database. A spatial pyramid based global image feature descriptor, a local feature matching PKD-trees and a refined MR-NLM scheme were presented. As future work, PKD-trees used for global feature vector, GPU acceleration for faster

execution and a more complete database will be tested.

REFERENCES

- [1] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, 490-530, 2005.
- [2] P. Coupe, P. Yger, and S. Prima et al., "An optimized blockwise non local means denoising filter for 3D magnetic resonance images," *Trans. on Med. Imag.*, vol. 27, 425-441, 2008.
- [3] G. Csurka, C. Bray, and L. Fan et al., "Visual categorization with bags of keypoints," In *ECCV Workshop on Stat. Learn. in Comp. Vis.* 2004.
- [4] T. Emrich, F. Graf, and H.-P. Kriegel et al., "CT Slice Localization via Instance-Based Regression," *Proc. SPIE*, vol. 7623, 76232O, 2010.
- [5] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE TSAP*, vol. 3, no. 6, 613-623, 1973.
- [6] X. Jia, Y. Lou, R. Li, W. Song, S. Jiang, "GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation," *Medical Physics*, vol. 37, pp. 3441-3447, 2010.
- [7] Z. Kelm, D. Blezek, B. Bartholmai, B. Erickson, "Optimizing non-local means for denoising low-dose CT," *IEEE Symp. on Biomedical Imaging (ISBI)*, pp. 662-665, 2009.
- [8] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," *Proc. CVPR*, vol. 2, 2169-78, 2006.
- [9] D. Lowe, "Object recognition from local scale-invariant features," *Inter. Conf. on Computer Vision*, pp. 1150-1157, 1999.
- [10] C. Liu, J. Yuen, A. Torralba, "SIFT flow: dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978-994, 2011.
- [11] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," *Visual Perception, Progress in Brain Research*, vol. 155, 2006.
- [12] E. Sidky, X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777-4807, 2008.
- [13] C. Silpa-Anan, R. Hartley, "Optimised KD-trees for fast image descriptor matching," *Proc. CVPR*, pp. 1-8, 2008.
- [14] T. Tasdizen, "Principal neighborhood dictionaries for Non-local Means image denoising," *Trans. Imag. Proc.*, vol. 18, no. 12, pp. 2649-60, 2009.
- [15] H. Yu, S. Zhao, E. Hoffman and G. Wang, "Ultra-low dose lung CT perfusion regularized by a previous scan," *Academic Radiology*, vol. 16, pp. 363-373, 2009.
- [16] Q. Xu, H. Yu and G. Wang et al., "Dictionary learning based low-dose X-ray CT reconstruction," *Proc. Fully3D*, pp. 258-261, Germany, 2011.
- [17] W. Xu, K. Mueller, "Evaluating popular non-linear image processing filters for their use in regularized iterative CT," *IEEE Medical Imaging Conference*, Knoxville, TN, October, 2010.
- [18] W. Xu, K. Mueller, "A reference image database approach for NLM filter-regularized CT reconstruction," *Proc. Fully3D*, pp. 116-119, 2011.
- [19] Boost C++ libraries, www.boost.org.

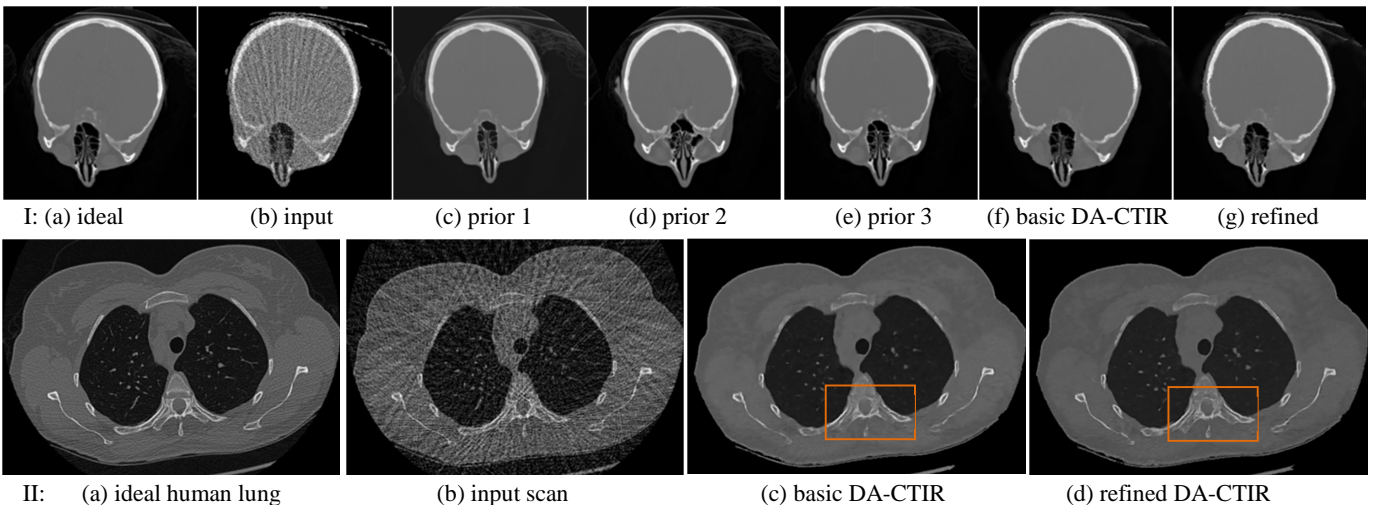


Figure 2: Results with (I) a CT head database and (II) a human lung database.