

Learning Effective Parameter Settings for Iterative CT Reconstruction Algorithms

Wei Xu and Klaus Mueller

Abstract—Iterative reconstruction algorithms are preferably used when the projection data are noisy, irregular in acquisition, or limited in number or size. They typically offer a set of parameters that allow some control over the convergence process, both in terms of quality and speed. Examples include relaxation factor, number of subsets, regularization coefficients, and the like. The interactions among these parameters and within the various data conditions can be complex, and thus effective combinations can be difficult to identify, leaving their choice often to educated guesses. We propose a data-driven learning approach to match given data configurations with their most effective reconstruction parameter configurations. We overcome the computational challenges associated with such a data-intensive approach by using commodity high-performance computing hardware (GPUs), which themselves have interacting parameters as well.

Index Terms—Iterative Algorithms, Ordered Subsets, Computed Tomography, GPU, Image Quality Metrics

I. INTRODUCTION

An effective way to limit the overall radiation dose a patient is subjected to in a CT scan is to reduce the number of projections and also the radiation per projection. Both, however, increase noise in the CT reconstructions, compromising contrast as well as resolution. Iterative algorithms have been shown to excel in these adverse settings, but due to the absence of an exact solution, careful parameter tuning is typically required to converge to a solution close to the exact. Examples for such parameters include relaxation factor, number of subsets, and regularization coefficients. Their choice is often made ad-hoc based on some prior experience, yet typically not endorsed by a certified level of confidence. An added difficulty is that parameters often interact in their effects on reconstruction speed and outcome. Thus it can be a non-trivial task to derive the most suitable combination for a given data scenario. There are two main strategies by which one may arrive at effective parameter settings: optimization and data-driven learning. Optimization is similar to the reconstruction process itself, seeking to find the optimal solution (here the parameter configuration) constrained by some objective function. However, optimization can be vulnerable to local minima and it also lacks in some sense the capability to adapt to new data scenarios. Learning, on the other hand, aims to determine a process model (described by parameters) from a set of collected observations. In our application, these observations are reconstructions obtained with parameterizations of a given iterative reconstruction algorithm, where the quality of the reconstructions then drives the parameterization.

Clearly, the more observations we can provide and the

Wei Xu and Klaus Mueller are with the Computer Science Department, Stony Brook University, Stony Brook, NY 11790 USA (phone: 631-632-1524; e-mail: {wxu, mueller}@cs.sunysb.edu). Funding was provided by NSF CCF-0702699, CCF-0621463 and NIH EB004099-01.

greater their diversity, the more accurate our model is set to be. An important factor in this context is the quality metric. Since we aim to provide reconstructions to be examined by human observers, we require a quality metric that is perceptually based. Further, since we strive for a large number of observations, this perceptual metric needs to be computer-based and efficient to compute. In the following, we present a metric fitting these requirements and then describe our reconstruction parameter-learning framework.

In our paper, Section 2 presents related work and background, Section 3 describes our parameter-learning framework, Section 4 presents results, and Section 5 ends with conclusions.

II. RELATED WORK AND BACKGROUND

Iterative methods can broadly be categorized into projection onto convex sets (POCS) algorithms (such as SART, SIRT, and POCS) and statistical algorithms (such as EM, OS-EM, and MAP). For the purpose of demonstrating the principles of our proposed approach, we select OS-SIRT (Ordered Subsets SIRT) [5]. OS-SIRT is a generalization of SART and SIRT, with SIRT having just one and SART having M subsets (with M being the number of projections). Its correction update is computed as:

$$v_j^{(k+1)} = v_j^{(k)} + \lambda \sum_{p_i \in OS_s} \frac{p_i - r_i}{\sum_{l=1}^N w_{il}} \quad r_i = \sum_{l=1}^N w_{il} \cdot v_l^{(k)} \quad (1)$$

Here, the weight factor w_{ij} determines the contribution of a voxel v_j to a ray r_i (starting from a projection pixel p_i in subset OS_s), and is typically given by the interpolation kernel. Hence, there are two parameters, the relaxation factor λ and the number of subsets S . Using these parameters the effects of noise and sparse views on reconstruction quality can be controlled, but their choice can also affect the number of iterations needed to converge. Typically, noisier projections require larger subset sizes (smaller S) and/or smaller λ -settings. On CPUs the subset size does not influence the speed of computation. On GPUs, however, an iteration with SART is typically the slowest due to the many projection-backprojection context switches which disturb parallelism and data flow. This has significant implications for the overall reconstruction wall clock time [5].

In CT reconstruction, the commonly used metrics for gauging reconstruction quality are mostly statistical, such as the cross-correlation coefficient (CC), root mean square (RMS) error, and R-factor. However, the assessment of image quality should include both objective and subjective metrics [1]. Objective metrics, such as blurriness and contrast, measure the physical and geometric properties of the image and their effect on human perception. Subjective methods, rooted in psychophysics, more formally introduce observer perceptual

metrics to gauge overall image quality. For example, Zhou et al. presented a comprehensive study of image comparison metrics [8], some described in the spatial and the frequency domain and some formulated in terms of the human visual system – the perception-based metrics – such as the visual differences predictor (VDP). While this (and many other) studies strongly suggest that perception-motivated metrics are superior to statistical ones, most of these involve heavy computation and thus are not desirable for the large data quantities we anticipate to process in our learning framework.

For few-view and noisy projection scenarios, the application of regularization operators between reconstruction iterations seeks to tune the final or intermediate results to some *a-priori* model. Total variance minimization (TVM) has commonly been used for noise and streak artifact reduction. But the iterative procedure of TVM is quite time-consuming, even when accelerated on GPUs. We therefore propose to use bilateral filtering [2], which is non-iterative, as an alternative regularization scheme. In a companion publication [7] we show that the bilateral filter can, in many scenarios, provide regularization effects of similar quality than TVM but at a fraction of its cost. The bilateral filter is a convolution operator which weighs a pixel neighborhood both in terms of closeness and similarity. We use Gaussian functions for these terms, adding two further parameters to our model:

$$c(\varepsilon, x) = e^{-\frac{\|x - \varepsilon\|^2}{2\sigma_d^2}} \quad s(\varepsilon, x) = e^{-\frac{(f(\varepsilon) - f(x))^2}{2\sigma_r^2}} \quad (2)$$

Here, σ_r and σ_d control the amount of smoothing and denoising.

III. METHODOLOGY

In the study presented here, in order to facilitate the computation of a ground truth-based figure of merit, we only used projection data acquired via simulation from known objects. We then added Gaussian noise at different SNR levels.

A. Overall Approach

Using the simulated projection data, we compute a representative set of reconstructions, sampling the parameter space in a comprehensive manner. We then evaluate these reconstructions with a perceptual quality metric, as discussed below. Adaptive sampling can be used to drive the data collection into more “interesting” parameter regions (those that produce more diverse reconstruction results in terms of the quality metrics). Having acquired these observations, we label them according to certain criteria, such as “quality, given a certain wall-clock time limit” or “reconstruction speed, given a certain quality threshold”. The observations with the higher marks, according to some grouping, subsequently receive higher weights in determining the reconstruction algorithm parameters. Currently, we either use the max-function or a fast-decaying Gaussian function to produce this weighting.

B. Image Quality Metrics

As mentioned, most popular for the assessment of image quality in CT have been statistical metrics such as mean absolute error (MAE), root mean square (RMS), normalized RMS (NRMS),

cross-correlation coefficient (CC) and R-factor. However, as also discussed above, these metrics do not consider the fact that human vision is highly sensitive to structural information [3]. These properties are well captured in the gradient domain, by ways of an edge-filtered image calculated via a Sobel Filter operator. We have labeled this group of metrics by prefix “E-”. For example, the E-CC metric stands for the CC of two edge images. Further, shifting effects caused by the CT reconstruction backprojection step can be alleviated by Gaussian-blurring the reconstruction image before edge-filtering. We have labeled this group of metrics with prefix “BE-”.

Another method to gauge structural information is Structural Similarity (SSIM) which combines luminance, contrast and structure [3]. Given two signal images x and y , the definition of SSIM index is defined as:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (3)$$

where α , β and γ are parameters adjusting relative importance. The terms $l(x, y)$, $c(x, y)$ and $s(x, y)$ are the luminance, contrast and structure comparison functions, respectively. These functions are computed from local image statistics [3].

IV. RESULTS

All computations used an NVIDIA GTX 280 GPU / Intel 2 Quad CPU 2.66GHz. We group our results into four sections: (1) the metrics comparison showing which metrics give consistent scores, (2) the OS-SIRT results showing the relationship between noise levels and parameters settings, (3) the OS-SIRT results for few-projection parameters settings, and (4) the parameter settings with our GPU-accelerated bilateral filter.

A. Comparing the Metrics

Our first (somewhat informal) study examined various image quality metrics and their suitability to replace a human judge/observer. We note that our emphasis in the onset was to select metrics that are fairly easy to evaluate, in order to keep the assessment of the expected large data volume manageable. We use a baby head scan (size 256^2) to demonstrate our results (see Fig. 1). All metrics are in the range $[0, 1]$ and thus allow direct numerical comparison. We compared the metrics at two levels: 0.99 (row 1 of Fig. 1) and 0.82 (row 2). We observe that the scores of E-CC and SSIM are quite faithful to the scores a

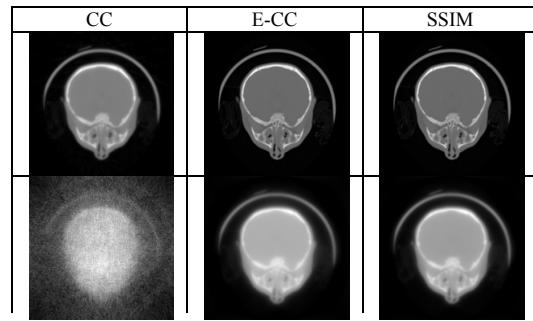


Fig. 1. Image metric comparison I: the value for each metric is fixed to 0.99 in the first row and to 0.82 in the second row.

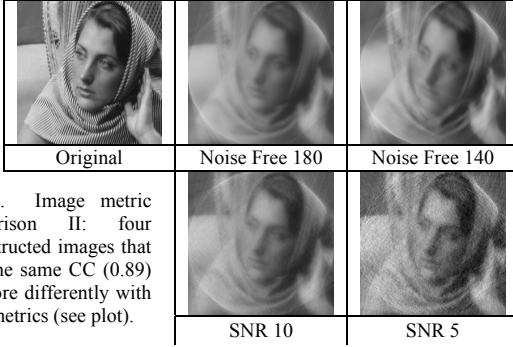


Fig. 2. Image metric comparison II: four reconstructed images that have the same CC (0.89) but score differently with other metrics (see plot).

human observer might give (0.99 for a near-perfect image and 0.82 for a “B+ similarity”). On the other hand, CC is clearly bound to over-score the images. It assigned the left bottom image a score of 0.82 even though the object structure is barely discernable, and it assigned the top left image a near-ideal score even though the structures are still quite blurry.

For a more comprehensive study of all metrics discussed in Section II, we reconstructed the popular “Barbara” test image (size 256^2), varying the number of projections and the level of noise (quantified by SNR). This image contains very high frequency detail and is thus quite sensitive to small errors. For all reconstructions we stopped iterations when the CC reached a value of 0.89. We chose this relatively low CC value so all SNR levels could reach it – better reconstructions are possible. Fig. 2 presents the images reconstructed from 180 and 140 noise-free projections (NF 180 and NF 140), respectively, and from 180 projections with Gaussian noise added (SNR 10 and SNR 5). We clearly see that these images do not look the same from a perceptual point of view, although the CC metric has the same outcome. Informally, an observer would likely rank these images in the order NF140, NF180, SNR10, SNR 5, with NF140 being the best (small detail is better visible there, although there are some slight yet tolerable artifacts). From the table in Fig 2 we observe that only the edge-based metrics (E-MAE, E-NRMS, and E-CC) as well as SSIM can reproduce this ranking (note that the ordering for CC is the opposite than for RMS and MAE since they have reverse maxima, and also note that we scaled some metrics by the given multiplicative factor to better visualize the contrast of the bars). The other metrics have either a wrong ranking or could not distinguish some images at all. In the remainder of the paper, we use E-CC since it better tolerates global density shifts and is faster to compute (than SSIM).

B. Learning Parameters

With a suitable quality metric in place we are now ready to learn the most effective parameters for the iterative OS-SIRT reconstruction algorithm we used to evaluate our approach. We first simulated, from the baby head CT scan, 180 projections at uniform angular spacing of $[-90^\circ, +90^\circ]$ in a parallel projection viewing geometry. We then added different levels of Gaussian noise to the projection data to obtain SNRs of 15, 10, 5, and 1. The first column of Fig. 4 presents the best reconstruction results (using the E-CC between original and reconstructed

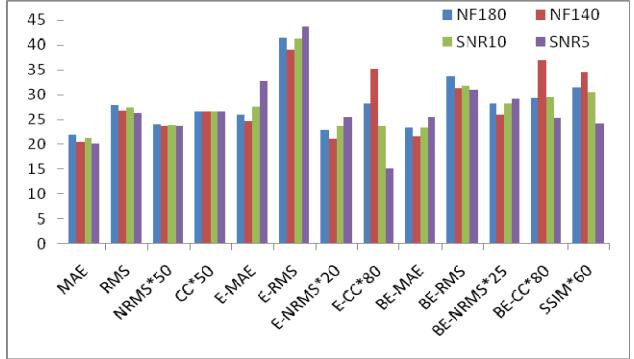


image), for each SNR, in terms of the “reconstruction speed, given a certain quality threshold” criterion. In other words, the images shown are the reconstructions that could be obtained at the shortest wall clock time given a certain minimal E-CC constraint. This constraint varies for each projection dataset (low SNR cannot reach high E-C levels), and this is also part of the process model.

Fig. 4 summarizes the various parameters obtained for the various data scenarios mentioned above. The “Best Subset” and “Best Lambda” values denote the parameter settings that promise to give the best results, in terms of the given quality metric and label criterion. The “Lowest Lambda” and “Turning Point” values describe the shape of the λ -curve as a function of the number of subsets. The λ -factor is always close to 1 for small subsets and then linearly (as an approximation) falls off at the “Turning Point” to value “Lowest Lambda” when each subset only consists of one projection (which is SART) [5].

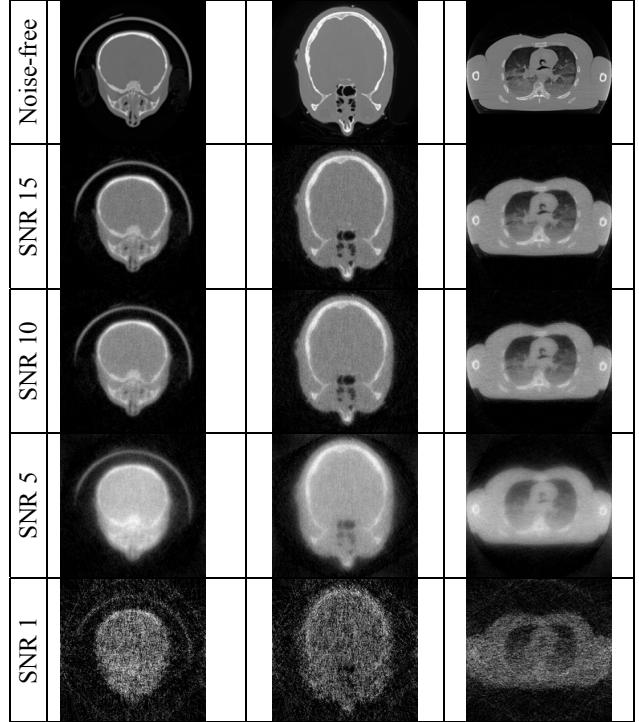


Fig. 3. Results for all SNR levels of three datasets using the same parameter settings, those found most effective for the baby head dataset: columns from left to right - baby head, visible human head, and visible human lung, rows from top to bottom - noise-free, SNR 15, 10, 5 and 1.

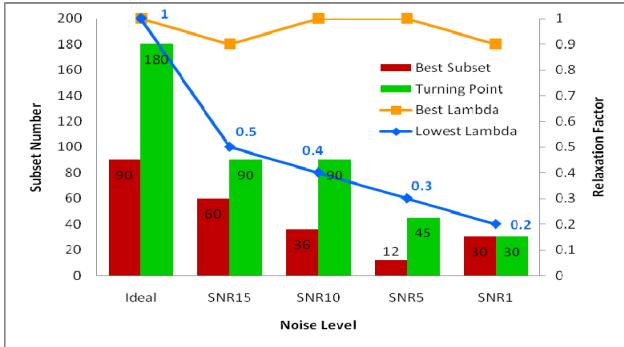


Fig. 4. Optimal parameter settings using E-CC for the baby head dataset: subset number and relaxation factor as a function of imaging condition (SNR) and the turning point and lowest lambda for each SNR level.

The summary plot of Fig. 4 helps practitioners to pick the best-performing number of subsets and the associated λ (to obtain the best possible quality within a given time) for a given expected SNR level. For example, we observe that low SNR requires a low number of subsets, while less noisy data can use a higher number of subsets. This trend is well confirmed by prior studies and field experience and thus validates the correctness of our general approach.

We then explored if the knowledge we learned translates to other similar data and reconstruction scenarios. Column 2 and 3 of Fig. 5 show the results obtained when applying the optimal settings learned from the baby head to reconstructions of the Visible Human head (size 256^3) and Visible Human lung (512^2), from similar projection data. We observe that the results are quite consistent with those obtained with the baby head, which is promising. As future work, we plan to compare the settings with those learned directly from these two candidate datasets.

Next, we investigated the few-view reconstruction scenario. Here, we “learned” that SART consistently gave the best results. Fig. 5 shows the best reconstructions of the baby head with $S=180$ to 20 in the top two rows. The third row shows the results of lung dataset using the same parameters than used in the second row. They are quite similar which confirms the generalization of the learned parameters.

Finally, we sought to learn the σ -parameters for the bilateral

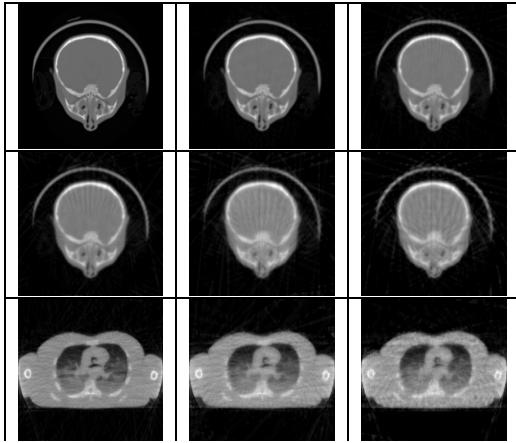


Fig. 5. Results for the few-view case: (first row) reconstructions from 180, 90 and 60 projections, (second row) reconstructions from 45, 30 and 20 projections, (third row) a lung reconstruction using the same parameters than the second row with 45, 30 and 20 projections.

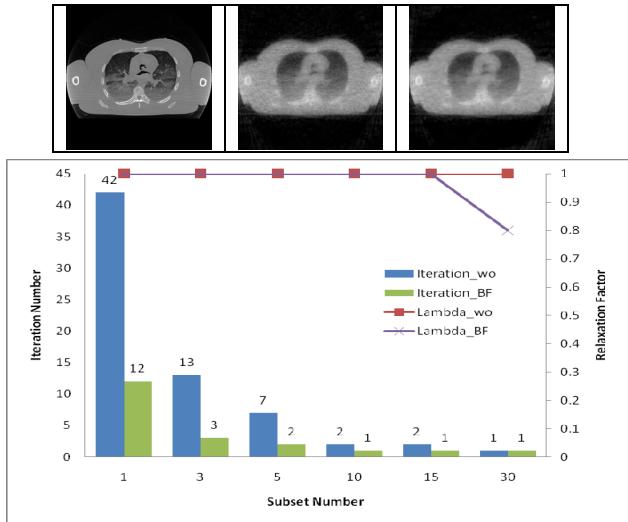


Fig. 6. (Top) The best results (using the time criterion) for the lung dataset with SNR 10 and 30 projections (E-CC=0.6). (left to right): original image, reconstruction without and with bilateral filter. (Bottom) Comparing the number of iterations required with / without bilateral-filter regularization.

filter used for regularization. Fig 6 presents results. We found that the bilateral filter helped to reduce the number of iterations especially for the smaller number of subsets. Since the bilateral filter is less expensive than an iteration it pays off to use it.

V. CONCLUSION

We demonstrated an intelligent framework that has the potential to automate the parameter selection for CT reconstruction tasks. Iterative algorithms likely benefit the most from this scheme, since they tend to have a variety of parameters to adapt the optimizer to the present data conditions and reconstruction goals. Such a system can be helpful to practitioners that do not have the expertise to tune these parameters by hand. Eventually, we hope to refine our framework such that it can recognize “signatures” directly from the projection data, combining them with other information about scanner and object, and use this information to index the parameter knowledge base.

REFERENCES

- [1] N. Birmingham, Z. Pizli, et al, “Image Quality metrics,” *Conf. Image Processing, Image Quality, Image Capture Systems*, pp. 598-616, 2003.
- [2] C. Tomasi, R. Manduchi, “Bilateral filtering for gray and color images,” *Intern. Conf. on Computer Vision (ICCV)*, pp. 839-846, 1998.
- [3] Z. Wang, A.C. Bovik, et al, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. on Img. Proc.*, 2004.
- [4] F. Xu, K. Mueller, “Accelerating popular tomographic reconstruction algorithms on commodity PC graphics hardware,” *IEEE Trans. on Nuclear Science*, 52: pp. 654-663, 2005.
- [5] F. Xu, K. Mueller, et al. “On the efficiency of iterative OS reconstruction algorithms for acceleration on GPUs,” *MICCAI Workshop on High-Performance Medical Image Computing*, NY, 2008.
- [6] F. Xu, K. Mueller, “Real-time 3D CT reconstruction using commodity graphics hardware,” *Phys. Med. Biol.*, 52: 3405–3419, 2007.
- [7] W. Xu, K. Mueller, “A Performance-Driven Study of Regularization Methods for GPU-Accelerated Iterative CT,” (submitted), *High Performance Image Reconstruction Workshop*, 2009.
- [8] H. Zhou, M. Chen, M.F. Webster, “Comparative Evaluation of Visualization and Experimental Results Using Image Comparison Metrics,” *IEEE Visualization*, pp. 315-322, 2002.