Fast and Efficient Stored Matrix Techniques for Optical Tomography

Guangzhi Cao, Charles A. Bouman and Kevin J. Webb
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907 USA
{gcao,bouman,webb}@ecn.purdue.edu

Abstract—A barrier to the use of optical tomography in practical applications is the high computational cost of iterative image reconstruction. This paper introduces a novel method for direct reconstruction of the image from a pre-computed and stored inverse matrix. Since the inverse matrix for optical tomography is generally quite large and not sparse, it is necessary to store the inverse matrix using lossy source coding techniques. A key innovation is the method used for matrix representation and the technique used for computing the required matrix-vector product. This representation is based on transforms of the image and sensor spaces which are designed to minimize reconstructed image distortion. Simulations indicate that the technique can dramatically reduce the storage and computation requirements by exploiting redundancy in the transformed matrix.

I. INTRODUCTION

Optical tomography methods such as optical diffusion tomography (ODT), fluorescence optical diffusion tomography (FODT), and bioluminescence tomography (BLT) potentially offer a safe, and inexpensive alternative to available tomographic technologies. However, due to multiple-scatting property of light in tissue, reconstruction in optical tomography is a highly nonlinear and ill-posed inverse problem. Therefore, reconstruction is typically formulated in an optimization framework, and iterative solution methods are used to obtain the reconstructed image. Although fast reconstruction algorithms have been developed [1]–[3], iterative reconstruction tends to be computationally expensive due to both the computation required for each iteration and the large number of iterations required for convergence. This computational burden is a potential barrier to the wide use of optical tomography in real-time clinical applications.

In this paper we present a novel non-iterative reconstruction approach for optical tomography with the potential to be much faster than iterative reconstruction methods. Our approach is to first pre-compute the linearized inverse system matrix for known values of the bulk homogeneous material parameters. The resulting inverse matrix is quite large and non-sparse, so it is necessary to use lossy compression methods to store it efficiently. The first step to lossy compression is to derive an appropriate distortion metric. We show that in order to achieve

minimum mean squared error in the reconstructed image, it is first necessary to apply a transformation to the rows of the inverse system matrix that decorrelates the measurement data. Next we propose the use of a Karhunnen-Loeve transform of the inverse-matrix rows, and a wavelet transform of the matrix columns since the matrix columns correspond to 3D reconstructed images. Finally, we use the SPIHT wavelet coding algorithm [4] to quantize and code the wavelet transformed matrix entrees.

The approach to lossy coding of the inverse matrix not only dramatically reduces storage, but it also reduces computation time since the number of entrees in the matrix which are effectively non-zero is dramatically reduced.

We present simulation results for the problem of FODT. For this problem, we show that the inverse system matrix can be compressed by factors of 200:1 with little effect on reconstruction quality. We also illustrate the reductions in computation speed and memory usage for a hypothetical example of an optical imaging surface probe in a FODT application.

II. INVERSE SYSTEM OF OPTICAL TOMOGRAPHY

First, we briefly review the model for FODT. The transport of continuous wave (CW) light in a fluorescent, highly scattering medium with an external source at the excitation wavelength is modelled by use of the coupled diffusion equations [5]:

$$\nabla \cdot [D_x(r)\nabla \phi_x(r)] - \mu_{a_x}(r)\phi_x(r) = -\delta(r - r_{s_k}) \quad (1)$$

$$\nabla \cdot [D_m(r)\nabla \phi_m(r)] - \mu_{a_m}(r)\phi_m(r) = -\phi_x(r)\eta\mu_{a_f}(r) \quad (2)$$

where the subscripts x and m, respectively, denote excitation and emission wavelengths λ_x and λ_m; $\phi(r)$ is the photon density; $\delta(r)$ the Dirac function; and r_{s_k} is the location of the kth excitation point source. The optical parameters are the diffusion coefficient $D(r)$ and the absorption coefficient $\mu_a(r)$, and $\eta\mu_{a_f}(r)$ is the fluorescence yield which incorporates the fluorophore's quantum efficiency η and its absorption coefficient $\mu_{a_f}(r)$.

Our goal is to reconstruct the fluorescence distribution $\eta\mu_{a_f}(r)$ using sources at λ_x and detectors filtered at λ_m. We use $y \in R^M$ to denote the surface measurements. After

discretizing the domain into N voxels of equal size, one can regard the unknown parameters as the image vector

$$x = [\eta\mu_{a_f}(r_1) \cdots \eta\mu_{a_f}(r_N)]^T . \tag{3}$$

Assuming that the bulk optical parameters of the medium μ_{a_x}, μ_{a_m}, D_x and D_m are known, the diffusion model can be used to determine the linear transformation $A \in R^{M \times N}$ from x to y. This results in the linearized forward model:

$$y = Ax \tag{4}$$

where $A_{ij} = \phi_x^{(i)}(r_j)g_m^{(i)}(r_j)$, g denotes the Green's function of the diffusion equation, and superscript i denotes the ith source-detector measurement pair.

In a Bayesian framework, the maximum *a posteriori* (MAP) estimation of x is given by

$$\hat{x} = \arg\max_{x \geq 0}\{\log p(y|x) + \log p(x)\} , \tag{5}$$

where $p(y|x)$ is the data likelihood and $p(x)$ is the prior model for image x. Assume that the noise is i.i.d. Gaussian and that x is a Gaussian Markov Random Field (GMRF). Then reconstruction can be formulated as the solution to the following optimization problem:

$$
\begin{aligned}
\hat{x} &= \arg\min_{x \geq 0} \parallel y - Ax \parallel_\Lambda^2 + \frac{1}{2\sigma^2}\sum_{(i,j)\in\mathcal{N}} b_{i-j}|x_i - x_j|^2 \\
&= \arg\min_{x \geq 0} \parallel y - Ax \parallel_\Lambda^2 + x^T S x ,
\end{aligned} \tag{6}
$$

where

$$S_{i,j} = \begin{cases} -b_{i-j}/4\sigma^2 & \text{if } i \neq j \\ 1/4\sigma^2 & \text{if } i = j \end{cases} .$$

Here, $\mathcal{N}$ is the set of all neighboring voxel pairs, σ controls the trade-off between fitting to the data and the smoothness of the solution, and b_{i-j} represents the coefficients assigned to neighbors i and j.

If we do not apply the positivity constraint, then we know the closed-form solution of (6) is given by

$$\hat{x} = (A^T\Lambda A + S)^{-1}A^T\Lambda y \tag{7}$$

Therefore, we can pre-compute the inverse system matrix

$$H \triangleq (A^T\Lambda A + S)^{-1}A^T\Lambda, \tag{8}$$

and store it. Then reconstruction results from multiplication of the inverse-system matrix H by the data vector y.

III. Optimal Lossy Compression for the Inverse System Matrix

In order to store the matrix H for practical problems, we will need to compress it using lossy compression methods. However, the performance of any lossy compression method depends critically on the distortion metric that is used. Therefore, we first must derive an appropriate distortion metric which minimizes mean-squared error in the reconstructed image. A good distortion metric should be a norm on the distortion of H that is proportional to the MSE of the image

$\hat{x}$. In order to obtain this metric, we must account for the distribution of the measurement y.

Let R_y be the auto-correlation matrix of y, i.e.,

$$
\begin{aligned}
R_y &\triangleq \mathrm{E}[yy^T] \\
&= E\Lambda_y E^T
\end{aligned} \tag{9}
$$

where E and Λ_y are the matrices of eigenvectors and eigenvalues of R_y, respectively. Define the transformed inverse system and data as

$$
\begin{aligned}
\tilde{H} &\triangleq HE\Lambda_y^{\frac{1}{2}} \tag{10} \\
\tilde{y} &\triangleq \Lambda_y^{-\frac{1}{2}}E^T y. \tag{11}
\end{aligned}
$$

Notice that $\mathrm{E}[\tilde{y}\tilde{y}^T] = I$, and $\hat{x} = \tilde{H}\tilde{y}$. If we further define δH as the quantization error in H and $\delta\hat{x}$ as the resulting quantization error in $\hat{x}$, then we have that

$$\delta\hat{x} = \delta\tilde{H}\tilde{y} . \tag{12}$$

From this we can obtain an expression for the conditional MSE in $\hat{x}$ given δH. Assume the measurement y is independent of the quantization error δH. Then we have the fact that

$$\mathrm{E}\left[\parallel \delta\hat{x} \parallel^2 | \delta H\right] = \parallel \delta\tilde{H} \parallel^2 = \parallel \delta H \parallel_{R_y}^2 \tag{13}$$

where $\parallel \delta\tilde{H} \parallel$ is the Frobenius norm, and

$$\parallel \delta H \parallel_{R_y}^2 = \mathrm{trace}\{\delta H^T R_y \delta H\}$$

The proof is given as follows.
Proof:

$$
\begin{aligned}
\mathrm{E}\left[\parallel \delta\hat{x} \parallel^2 | \delta H\right] &= \mathrm{E}\left[\tilde{y}^T\delta\tilde{H}^T\delta\tilde{H}\tilde{y} \mid \delta\tilde{H}\right] \tag{14} \\
&= \mathrm{E}\left[\mathrm{trace}\{\tilde{y}^T\delta\tilde{H}^T\delta\tilde{H}\tilde{y}\} \mid \delta\tilde{H}\right] \\
&= \mathrm{E}\left[\mathrm{trace}\{\delta\tilde{H}\tilde{y}\tilde{y}^T\delta\tilde{H}^T\} \mid \delta\tilde{H}\right] \\
&= \mathrm{trace}\{\delta\tilde{H}\mathrm{E}[\tilde{y}\tilde{y}^T]\delta\tilde{H}^T\} \\
&= \mathrm{trace}\{\delta\tilde{H}\delta\tilde{H}^T\} \\
&= \parallel \delta\tilde{H} \parallel^2 .
\end{aligned}
$$

∎

This means that if we minimize $\parallel \delta\tilde{H} \parallel^2$ we obtain a reconstructed image $\hat{x} + \delta\hat{x}$ with minimum distortion. Therefore, $\tilde{H}$ is an appropriate representation for the storage and coding of the inverse system matrix H. Actually $\tilde{H}$ can be computed directly instead of H. Specifically, the ith column of $\tilde{H}$ can be computed as below:

$$\hat{\tilde{h}}_i = \arg\min_x \parallel \tilde{e}_i - Ax \parallel_\Lambda^2 + x^T S x , \tag{15}$$

where $\tilde{e}_i = E\Lambda_y^{1/2}e_i$. Here e_i denotes the ith unit vector.

Using this distortion metric, our next goal is to find a sparse representation for $\tilde{H}$ and to use the sparse representation for efficient coding and storage. Like H, $\tilde{H}$ is a $N \times M$ matrix whose columns are 3D images corresponding to different sensor measurements. Therefore, we can decorrelate along rows with Karhunen-Loeve (KL) transform (which results in
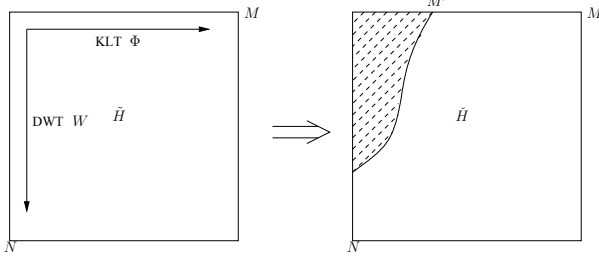
Fig. 1. Illustration of inverse system matrix compression. $\check{H}$ is a sparse representation of $\tilde{H}$ through KL and wavelet transforms. The shaded region represents the effective non-zero entrees in $\check{H}$.

eigen-images) and decorrelate along columns with wavelet transform [6]. These transforms can be expressed as follows:

$$\check{H} = W\tilde{H}\Phi \tag{16}$$

where Φ consists of the eigenvectors of $\tilde{H}^T\tilde{H}$, and W represents 3D wavelet transform. Fig. 1 illustrates the representation of the matrix $\check{H}$. Since both the KL and wavelet transform are orthonormal, we know $\| \delta\check{H} \|^2 = \| \delta\tilde{H} \|^2$. Accordingly, define the transform data as

$$\check{y} \stackrel{\triangle}{=} \Phi^T\tilde{y} = \Phi^T\Lambda_y^{-\frac{1}{2}}E^Ty \ . \tag{17}$$

Then $\hat{x}$ can be expressed as:

$$\hat{x} = W^{-1}\check{H}\check{y} \tag{18}$$

After obtaining the sparse representation $\check{H}$, Set Partitioning In Hierarchical Trees algorithm (SPIHT) [4] is used to quantize and encode each column of the matrix. Note we also need to correctly implement bit-allocation among different columns. Since the KL transform is an orthonormal linear transformation, the wavelet coefficients in each column should be quantized using the same quantization table. This is realized by explicitly specifying the same bit-planes in SPIHT encoding of each column.

Fig. 1 illustrates compression of $\tilde{H}$, where $\check{H}$ is the sparse representation of $\tilde{H}$ obtained through KL and wavelet transforms. Due to sparsity of $\check{H}$, evaluation of $\check{H}\check{y}$ requires many fewer multiplications than evaluation of Hy. Therefore, lossy compression both reduces the data that must be stored, and reduces computation required to reconstruct the image.

IV. COMPUTATIONAL COMPLEXITY

Here, let us analyze the computational complexity of the proposed algorithm for reconstruction. First define

$$T \stackrel{\triangle}{=} \Phi^T\Lambda_y^{-\frac{1}{2}}E^T. \tag{19}$$

Then from (18), reconstruction can be rewritten as follows:

$$\hat{x} = W^{-1}[\check{H}]Ty \tag{20}$$

where $[\check{H}]$ denotes the quantized version of $\check{H}$. The computational complexity of (20) is straightforward: MM multiplications are required for calculating Ty, and cNM/r for the product to multiply $[\check{H}]$, where r is the compression ratio and c/r

gives the ratio of non-zero coefficients after quantization. An additional order N multiplications are required for the inverse wavelet transform, which is small compared to the previous two terms. Therefore, the total computational complexity of the proposed algorithm is $cNM/r + MM$. The memory usage of the algorithm is as follows: order cNM/r for $[\check{H}]$ and order N for the reconstructed image $\hat{x}$.

Note the transform matrix $T^{M \times M}$ also needs to be stored as side information. In optical tomography, the size of T is small compared to that of $\check{H}$ since typically the number of measurements M is much less than the number of voxels N. Therefore, the total memory requirement is $cNM/r + MM$. Comparison of the computational complexity based on the CG method using a stored matrix A and the proposed new reconstruction method for FODT is given in Table I along with a numerical example which corresponds to the simulation example in next section. From the table, we can clearly see the advantage of the new algorithm over CG both in terms of computational complexity and memory requirement.

V. NUMERICAL RESULTS

Here we give some numerical results of FODT from reflectance measurements based on the proposed algorithm. Reflectance measurement system is clinically appealing, however, it also provides a very challenging tomography problem because it is much more ill-conditioned than in the case of the transmission measurement geometry.

The measurement geometry is shown in Fig. 2(a) where a 6×6 cm^2 probe scans on top of the semi-infinite medium. The probe contains 4 sources and 117 detectors as shown in Fig. 2(b). The background optical values are set to $\mu_{a_{x,m}} = 0.02$ cm^{-1}, $D_{x,m} = 0.03$ cm and $\eta\mu_{a_f} = 0$ cm^{-1}, which approximates the optical property of tissue. The measurements were generated with a spherical heterogeneity of radius 0.5 cm present 2 cm below the center of the probe. The optical values of the heterogeneity are $\mu_{a_x} = 0.12$ cm^{-1}, $\mu_{a_m} = 0.02$ cm^{-1}, $D_x = D_m = 0.03$ cm^{-1} and $\eta\mu_{a_f} = 0.05$ cm^{-1}. MUDPACK [7] was used as our forward PDE solver. The size of the computation domain is $8 \times 8 \times 4$ cm^3 with a resolution 0.25 cm. Additive noise was introduced based on the shot noise model of [2] giving an average SNR of 35.6 dB. The forward system A was computed using the bulk optical values assuming homogeneous medium. Use of the reciprocity of the Green function in the diffusion equation can dramatically reduce the computation of A.

Since the covariance matrix of the shot-noise model is a function of the measurements, we are not able to incorporate it into our pre-computation of the inverse system. Instead, the average noise variance is used in the noise model for pre-computation of the inverse system. This is currently a limitation of our algorithm. But if SNR is not too low, reconstruction results based on the two different noise models are usually quite close. In order to obtain better resolution in the deep medium, the regularization parameter σ was linearly increased with the depth. To model the statistics of the measurements, we have $R_y = AR_xA^T$. $R_x = I$ was

TABLE I

COMPUTATIONAL COMPLEXITY ANALYSIS FOR FODT. $N = 33 \times 33 \times 17$ - NUMBER OF VOXELS, $M = 468$ - NUMBER OF MEASUREMENTS, $I = 100$ - NUMBER OF ITERATIONS, $r = 192$ - COMPRESSION RATIO AND $c = 6.7$.

	computational complexity			Memory Usage		
	Order	example	CPU time	Order	example	RAM usage
CG	$2MNI$	1.7×10^9	70sec	MN	8.6×10^6	70MB
New method	$cNM/r + M^2$	5.2×10^5	0.015sec	$cNM/r + M^2$	5.2×10^5	4.1MB
Uncompressed	NM	8.6×10^6	0.25sec	NM	8.6×10^6	70MB

(a) measurement system

(b) Probe configuration: × - source; ○ - detector
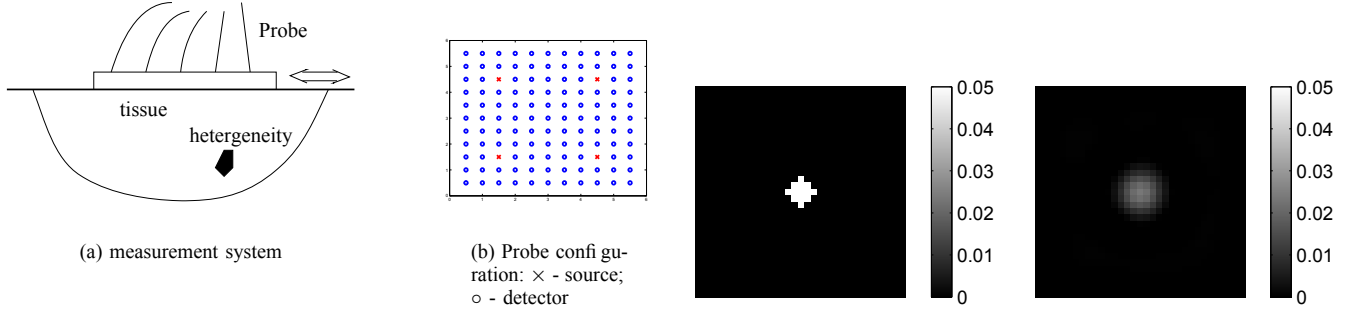
Fig. 2. Reflectance measurement geometry

(a) Original image

(b) 1:1

taken in simulations. The wavelet transform was constructed with the popular biorthogonal 9/7 tap filters [8] and using a symmetric boundary extension. We used QccPack library [9] for implementing 3D wavelet SPIHT coding.

Fig. 3 shows the reconstruction results at the depth of 2 cm with the stored inverse matrix compressed at different ratios. The distortion versus compression ratio curve is given in Fig. 4, where the distortion is calculated in terms of the normalized root mean squared error (NRMSE) defined as:

$$\text{NRMSE} = \sqrt{\frac{\| [H]y - Hy \|^2}{\| Hy \|^2}}. \tag{21}$$

The distortion-compression curve using Frobeneius norm $\| \delta H \|^2$ as the distortion metric is also given. For the same distortion, twice the compression ratio can be achieved using the correct distortion metric. The simulations were run in a 64bit dual processor Intel machine. Both CPU time and RAM usage at the compression ratio 192:1 are given in Table I along with a comparison to both CG and reconstruction using the uncompressed inverse matrix. We can see that compression of the inverse matrix both reduces the data storage and computation effectively.

VI. CONCLUSION

In this paper we presented a novel reconstruction approach for FODT using the stored matrix techniques. Compared to conventional iterative reconstruction algorithms, our new method offers faster and more efficient reconstruction both in terms of computational complexity and memory usage. This makes the new method very attractive to real applications. A method for lossy compression of the inverse matrix is also provided. Numerical simulations show that compression of the

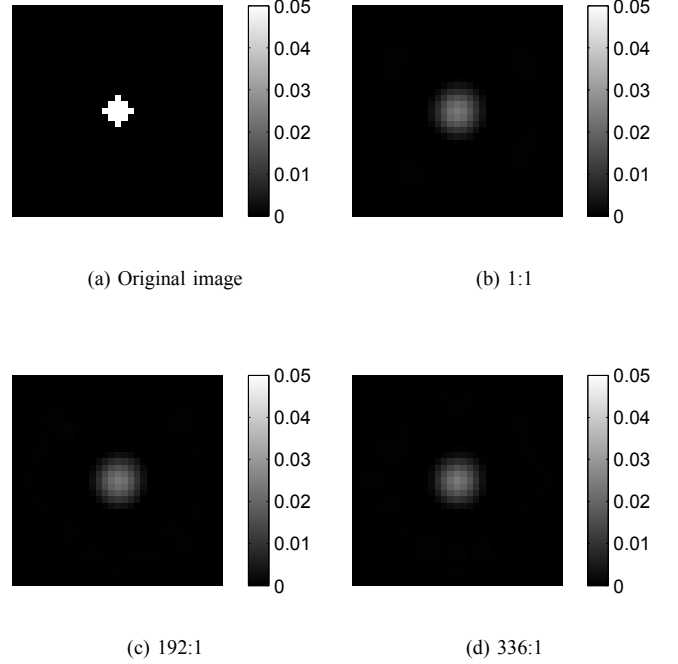

(c) 192:1

(d) 336:1

Fig. 3. Reconstruction results at different compression ratios
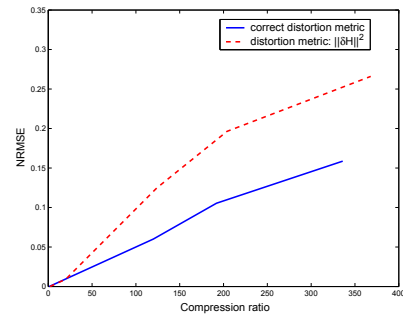

Fig. 4. Distortion versus compression ratio

inverse matrix can be quite high, which in turn leads to more efficient computation of the matrix-vector product required for reconstruction. The same approach can be extended to other optical tomography methodologies, e.g. ODT and BLT. For future work, we would like to extend our compression method to the forward model to improve the computation speed in an iterative reconstruction framework. Also, we would like to extend our approach to more general tomography methodologies by addressing the problem when number of measurements is large.

References

[1] S. R. Arridge and M. Schweiger, "A gradient-based optimisation scheme for optical tomography," *Optics Express*, vol. 2, no. 6, pp. 213–226, March 1998.

[2] J. C. Ye, K. J. Webb, C. A. Bouman, and R. P. Millane, "Optical diffusion tomography using iterative coordinate descent optimization in a Bayesian framework," *J. Optical Society America A*, vol. 16, no. 10, pp. 2400–2412, October 1999.

[3] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb, "A general framework for nonlinear multigrid inversion," *IEEE Trans. on Image Processing*, vol. 14, no. 1, pp. 125–140, Jan. 2005.

[4] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.

[5] A. B. Milstein, S. Oh, K. J. Webb, C. A. Bouman, Q. Zhang, D. A. Boas, and R. P. Millane, "Fluorescence optical diffusion tomography," *Applied Optics*, vol. 42, no. 16, pp. 3081–3094, June 2003.

[6] D. Tretter and C. A. Bouman, "Optimal transforms for multispectral and multilayer image coding," *IEEE Trans. on Image Processing*, vol. 4, no. 3, pp. 296–308, March 1995.

[7] J. C. Adams, "MUDPACK: Multigrid portable FORTRAN software for the efficient solution of linear elliptic partial differential equations," *Appl. Math. Comput.*, vol. 34, pp. 113–146, 1989.

[8] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Processing*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[9] J. E. Fowler, "Qccpack: An open-source software library for quantization, compression, and coding," in *DCC '00: Proceedings of the Conference on Data Compression*. Washington, DC, USA: IEEE Computer Society, 2000, p. 554.