A hybrid approach for diffuse optical tomography combining evolution strategies and gradient techniques

Alexander D. Klose [*a,c] and Andreas H. Hielscher [**a,b]

[a] State University of New York - Downstate Medical Center, Brooklyn, NY
[b]Dept. of Electrical and Computer Engineering, Polytechnic University, Brooklyn, NY
[c] Freie Universität Berlin, Inst. Medizinische Physik, Berlin, Germany

ABSTRACT

Diffuse optical tomography (DOT) can be considered as an optimization problem, in which the minimum of an objective function is sought. The objective function is typically some measure of the difference between the predicted and experimentally obtained detector readings. Most of the optimization techniques that are currently applied in optical tomography employ so-called gradient methods. These methods start from an initial guess of the distribution of optical properties and iteratively update this initial guess along the gradient of the objective function. It is well known that the success of gradient techniques depends strongly on the initial guess. If the guess is not chosen appropriately, the algorithm may not converge or may converge to a local minimum. Evolution strategies are global optimization techniques that depend much less on initial guesses. The drawback of evolution-based codes is that they are computationally expensive. In this work we introduce a hybrid approach that combines the advantages of gradient techniques and evolution strategies. The hybrid algorithm is less dependent on an initial guess and overcomes the computational burden connected to evolution strategies.

Keywords: optical tomography, image reconstruction, scattering media, global optimization, conjugate gradient technique, evolution strategy, transport theory.

1. INTRODUCTION

The most promising reconstruction algorithms for diffuse optical tomography (OT) are so-called model-based iterative image reconstruction schemes (MOBIIR).[1-8] In these schemes the image reconstruction problem can be formulated as a numerical optimization problem, which consists of three major parts. First, a forward model for light transport is used to predict the measured data. The input to this forward model is the spatial distribution of optical properties, such as the scattering coefficient μ_s and the absorption coefficient μ_a, and source positions. Second, an objective function, often also referred to as error function, is evaluated to obtain a measure for the difference between predicted and measured data. In a third step the initial guess of the spatial distribution of optical properties is updated in such a way that the value of the error function is reduced. These steps are repeated until the error function is minimized.

The update of the spatial distribution of optical properties is often obtained by using the gradient of the error function with respect to the optical parameters. However, it is well known that such gradient-based updating schemes are very sensitive to the initial guess, and often get trapped in a local minimum close to the initial guess.

Recently, our group introduced the concept of evolution-strategy-based optimization for optical tomography.[9] In general, evolution strategies mimic the evolution process in nature to find the global minimum in optimization problems. The disadvantage of this approach is that the computational burden growth quickly with the number of unknowns. Therefore, we have developed a hybrid approach, which combines the advantages of a computationally efficient gradient scheme with a global-minimum-finding evolution strategy. This method uses an evolution strategy to find a good initial guess for the gradient-based scheme. Using numerical simulations, we found that this approach considerably improves the image quality of our reconstructions.

[*]aklose@downstate.edu; [**]ahielscher@downstate.edu; State University of New York - Downstate Medical Center, Dept. of Pathology, 450 Clarkson Avenue, Box 25, Brooklyn, NY 11203

Optical Tomography and Spectroscopy of Tissue IV, Britton Chance, Robert R. Alfano, Bruce J. Tromberg, Mamoru Tamura, Eva M. Sevick-Muraca, Editors, Proceedings of SPIE Vol. 4250 (2001)
© 2001 SPIE · 1605-7422/01/$15.00

Downloaded from SPIE Digital Library on 23 Nov 2009 to 128.59.151.11. Terms of Use: http://spiedl.org/terms

2. METHODS

2.1 Model-based iterative image reconstruction

Using model-based iterative image reconstruction schemes, the imaging problem in DOT can be interpreted as an optimization problem, where an optimization method seeks the minimum of an appropriately defined objective function. In general, the objective function determines the difference of the predicted and measured detector readings. In this work we define the objective function as the χ^2-error norm, which is given by

$$\Phi = \sum_s \sum_d \frac{\left(M_{s,d} - P_{s,d}(\mu_a, \mu_s)\right)^2}{{M_{s,d}}^2}. \quad (1)$$

Here, $M_{s,d}$ are the measured data and $P_{s,d}$ are the predicted detector readings for a given source position s and detector position d. The predicted detector readings are obtained with a forward model for light propagation in turbid media. In general the forward model, and therefore the objective function, depends nonlinearly on the spatial distribution of the optical parameters μ_a and μ_s. Most of the currently employed forward models are based on the diffusion approximation to the equation of radiative transfer (ERT).[8] However, in this work we use the time-independent ERT.

$$\omega \cdot \nabla\Psi(r,\omega) + \left(\mu_a + \mu_s\right)\Psi(r,\omega) = S(r,\omega) + \mu_s \int_0^{2\pi} p(\omega,\omega')\Psi(r,\omega')d\omega'. \quad (2)$$

The fundamental quantity in radiative transport theory is the radiance $\Psi(r,\Omega)$ at the spatial position r, which is directed into a unit solid angle Ω. The radiance has units of [W cm^{-2} sr^{-1}], and the scattering and absorption coefficients have both units of [cm^{-1}]. The phase function $p(\omega,\omega')$ describes the scattering characteristics of the medium. A commonly used phase function is the Henyey-Greenstein function, which is defined as [10]

$$p(\cos\theta) = \frac{1-g^2}{2(1+g^2-2g\cos\theta)^{1.5}}. \quad (3)$$

Here θ is the scattering angle, and g is the anisotropy factor, which is equal to 0 for isotropically scattering media, $-1 < g < 0$ for backscattering media, and $0 < g < 1$ for forward scattering media. The reduced scattering coefficient μ_s' is defined by $\mu_s' = (1 - g)\,\mu_s$. The boundary conditions of the transport equation are described by Fresnel's law for interfaces of different refraction indices. We used a finite-difference discrete-ordinate method to solve the ERT for the radiance $\Psi(r,\Omega)$. Details of this approach are described elsewhere.[5]

A simple example of an objective function is shown in Fig. 1, where the value of Φ(Eq. 1) is plotted as a function of μ_s and μ_a. The graph was obtained in the following way: We assumed a 4cm x 4cm homogeneous medium with $\mu_s = 6.0$ cm^{-1}, μ_a=0.6 cm^{-1}, and g = 0.5. Placing one source in the center of each side, we use our forward model to calculate the outward pointing radiance for sixty detector positions that were placed in equal distances around the medium. Since 4 sources were used, four forward calculations were necessary to generate the full data set of detector readings $M_{s,d}$. This simulated measurement data was furthermore corrupted with 30dB Gaussian noise. Once these "measured" data was generated, predictions $P_{s,d}$ were calculated for various sets of optical parameters (μ_s, μ_a, g = 0.5). Each point of the objective function pertains to a different set of optical parameters (μ_s,μ_a). The objective function has a minimum at (μ_s= 6.0 cm^{-1},μ_a= 0.6 cm^{-1}), where the predicted detector readings match the "measured" detector readings.

Downloaded from SPIE Digital Library on 23 Nov 2009 to 128.59.151.11. Terms of Use: http://spiedl.org/terms

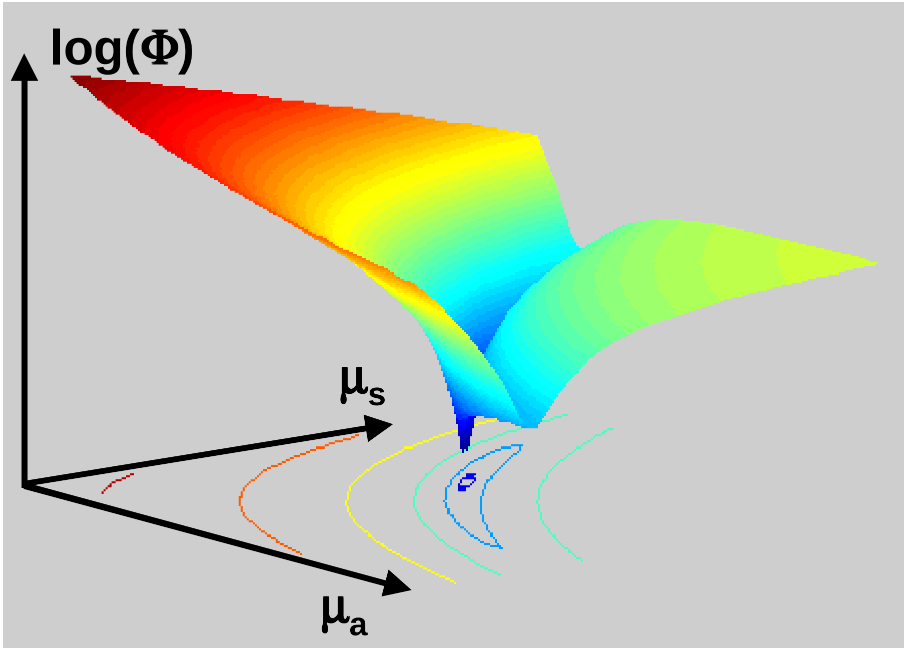

Figure 1. Example of an objective function Φ(Eq. 1) for a 4cm x 4cm homogeneous medium with $\mu_s = 6.0$ cm^{-1}, μ_a=0.6 cm^{-1}, and g = 0.5.

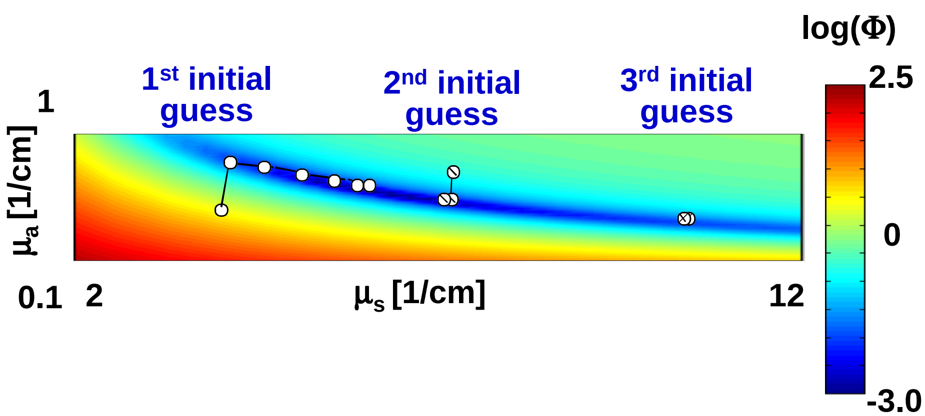

Figure 2. Same objective function as in Fig. 1. Shown are the iterative updates of optical properties (μ_s,μ_a) for three different initial guesses, using a conjugate-gradient updating scheme. Only when started with initial guess #1, the algorithm does find the minimum at ($\mu_s = 6.0$ cm^{-1}, μ_a=0.6 cm). The other initial guesses lead to reconstruction values that are inside the long narrow valley of (μ_s,μ_a) pairs that have similar small values of Φ.

2.2 Gradient-Based Optimization

In general, MOBIIR schemes seek the minimum of the objective function by starting from an initial guess of the spatial distribution of optical properties. Most of the currently available algorithms are using some form of gradient-based optimization scheme for this task.[11] The gradient $\partial\Phi/\partial(\mu_a,\mu_s)$ of the objective function with respect to the optical properties can be obtained by various techniques. In this work we use an adjoint differentiation method, which has been embraced by an increasing number of groups over the recent years. Once the gradient is obtained iterative line-minimizations along the gradient of the objective function are performed starting from an initial guess.[11]

Three examples of this iterative search procedure are shown in Fig. 2. We see that for each initial guess the algorithm stops at a different set of optical properties. Only the first initial guess located in the lower left corner of the search space leads to

Downloaded from SPIE Digital Library on 23 Nov 2009 to 128.59.151.11. Terms of Use: http://spiedl.org/terms

the global minimum. Starting from the other initial guesses the gradient-based scheme "gets stuck" in a location inside the valley far away from the minimum. Figure 3 displays the objective function along the bottom of the valley as a function of μ_s. The value of the objective function for the initial guess is indicated by a black dot. The value of the objective function for the reconstructed (μ_s, μ_a) pair is represented by a white dot. The gradient within the valley is steeper to the left of the minimum than to the right. This explains, why only starting from the 1st initial guess the minimum is found. The steeper the gradient, the better the gradient based algorithm will perform.

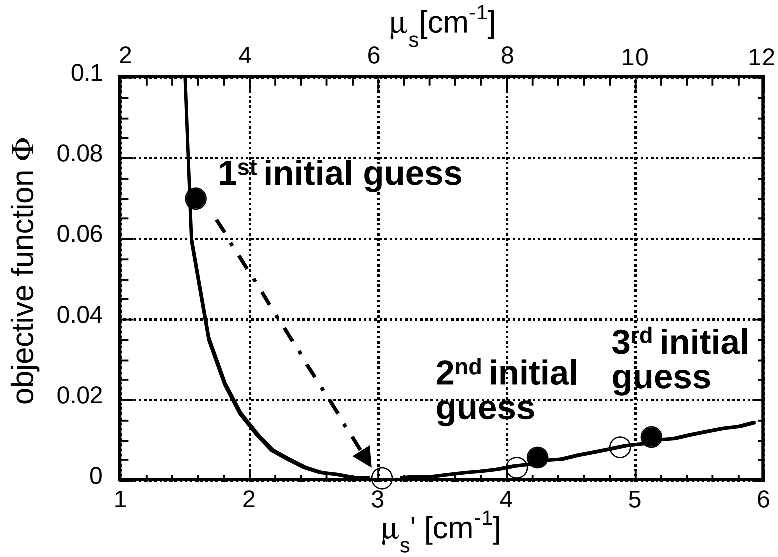

Figure 3. Values of the objective function at the bottom of the elongated, curved valley in Fig. 2. It can be seen that the gradient of the objective function is much steeper to the left of the minimum at μ_s = 6 cm^{-1} than to the right of the minimum. The line-minimizations along the valley of the objective function started from different initial guesses (black dots). Only starting from the 1st initial guess leads to the minimum.

2.3 Evolution Strategies

Evolution strategies are global optimization schemes, which do not use a gradient to find the minimum.[9,12-14] The general concept of the evolution strategies is adapted from the evolutionary process in nature. Each member of a population is represented by its phenotype. The phenotype consists of variables that represent the characteristics of each individual of the population. These

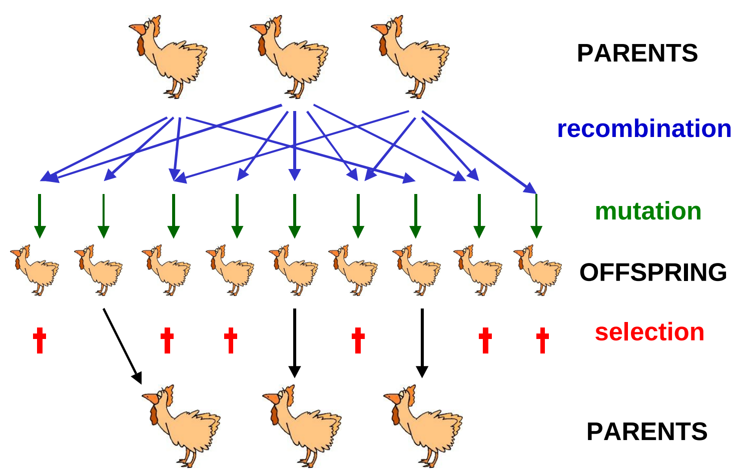

Figure 4. One generation cycle of the evolution process in nature. A population of three parents produces offspring by recombination and mutation The fittest offspring become the new parent generation by the process of selection.

Downloaded from SPIE Digital Library on 23 Nov 2009 to 128.59.151.11. Terms of Use: http://spiedl.org/terms

characteristics are changed by the processes of recombination, mutation and selection. Recombination is the first operation during one generation cycle. Pairs of individuals of the current population are randomly chosen to become parents, who interchange their variables of the phenotype. The variables are further modified within the mutation process. The final result of these two steps is the offspring. The fittest offspring survive the selection process to become the parents of the next generation cycle (see Fig. 4). This process is repeated for many generation cycles until the parents will be best adapted to their environment.

To illustrate the concept for DOT we have focused on the example we used for the gradient-based optimization in Fig. 2. In this case we deal with only two unknowns, $\boldsymbol{\xi} = (\xi_1 = \mu_a, \xi_2 = \mu_s)$. We start the search process by randomly choosing three different $\boldsymbol{\xi}$ with values of $0.1 < \mu_a < 1$, $2.0 < \mu_s < 12$, and g = 0.5. This is our parent population. The vectors $(\xi_1, \xi_2)^k = (\mu_a, \mu_s)^k$ with k=1,2,3, represent the phenotypes of each member of the parent population.

Next we need to generate the offspring that correspond to new vectors $\boldsymbol{\xi}$.. In our case, this is done by randomly pairing 12 couples from the parent population. The following six different pairs of parents are equally likely to be chosen: $[\xi^1\text{-}\xi^1]$, $[\xi^1\text{-}\xi^2]$, $[\xi^1\text{-}\xi^3]$, $[\xi^2\text{-}\xi^2]$, $[\xi^2\text{-}\xi^3]$, $[\xi^3\text{-}\xi^3]$. From each pair of parents one child is generated in a two-step process. In the first step properties of the two parents are recombined. In our case we chose a simple recombination rule by just averaging the optical properties of both parents:

$$\xi_i^{new*} = (\xi_i^{k_1} + \xi_i^{k_2}) / 2 . \tag{4}$$

In the second step, the average optical properties are further modified by the mutation process. We define the mutation rule as

$$\xi_i^{new} = \xi_i^{new*} + \sigma_i y \quad \text{for i} \in \{1,2,3\}, \tag{5}$$

where y is a normally (Gaussian) distributed random number, which is obtained from

$$y = \sqrt{-2\log(v_1)}\,\sin(2\pi v_2) \tag{6}$$

with v_1 and v_2 being independently and uniformly distributed between 0 and 1. The values of σ_i where set to $\sigma_i = 1/3\ \xi_{i,old}$. Therefore, the optical properties of the offspring are Gaussian distributed around the average optical properties of the parents. The width of the Gaussian distribution is given by 1/3 of the actual value of μ_a, μ_s. Other values for σ and other distributions are also possible but have not been explored by us at this point.[13]

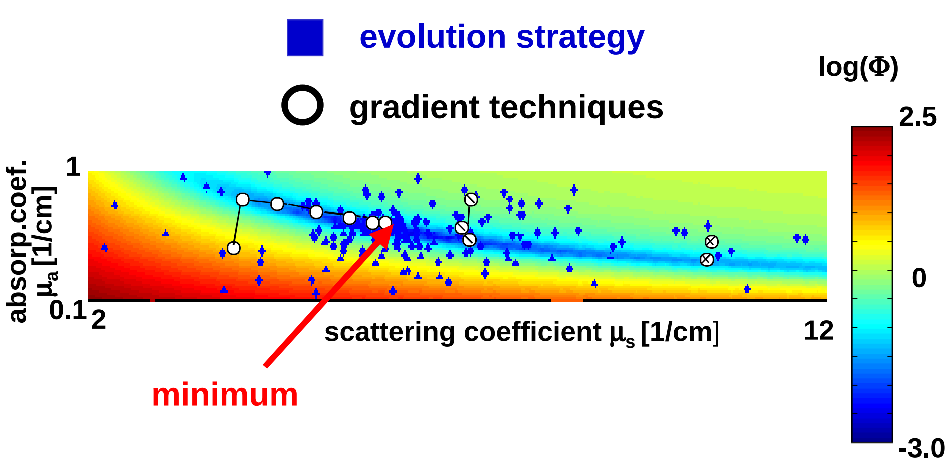

Figure 5. Comparison of the evolution strategy with the conjugate gradient technique.

Once the offspring is generated, the fitness off all members of the offspring is evaluated to select the three fittest members as the new parent generation. The selection is done by performing a set of forward calculations using the finite-difference discrete-ordinate transport code. The input parameters for the forward code are source and detector positions, and the optical properties $(\mu_a, \mu_s)_i$ that characterize each member. The results of the forward calculation are used to determine the value of

Downloaded from SPIE Digital Library on 23 Nov 2009 to 128.59.151.11. Terms of Use: http://spiedl.org/terms

the objective function as given in Eq. 1. The lower the value of the objective function, the "fitter" is that member of the offspring population. Only the three sets of optical properties (children) that result in the lowest values of the objective function survive, and become the new parents for the next generation. This procedure is iterated until the optical properties of the next generations converge within 2%. The results are shown in Fig. 5.

2.4 Hybrid Approach

Applying evolution-strategy-based schemes to heterogeneous media, which contain a larger number of unknowns, will result in computationally more intensive operations than the simple homogeneous example discussed so far. For example the convergence speed of an evolution strategy is commonly described by the rate of progress φ. The rate of progress is the ratio of the number of successful mutations to the total number of mutations. Using Rechenberg's sphere model the maximum rate of progress φ_{max} is inversely proportional to the number of unknown variables n with $\varphi_{max} \propto 1/n$ [12, pages 127-141]. That means if the number of variables is increased the maximum rate of progress is decreased, which leads to fewer successful mutations and therefore to a lower convergence speed. In general that means that more generations are required, which in turn lead to a larger number of forward calculations. Furthermore, larger populations are necessary, which will further slow down the optimization process. On the other hand, the gradient technique is more suitable for problems with many unknowns. But as seen in the previous sections, gradient techniques are highly dependent on the initial guess.

In this work we developed a hybrid approach that combines the gradient technique and the evolution strategy. The evolution strategy is used to find a good initial guess for the gradient-based optimization technique. Unlike the final image, which typically has more than 1000 pixels (unknown optical properties), we describe the initial guess with a limited number of parameters. The subspace of limited number of parameters is optimized using the evolution strategy. Therefore, the space of unknowns is initially randomly sampled to find a set of parents. Starting from these randomly selected samples (parents) subsequent generations are derived using recombination, mutation and selection schemes. Once the global minimum for this subset of optical parameters is found, the gradient-based optimization scheme is used to further minimize the objective function by allowing variations in all pixels of the image.

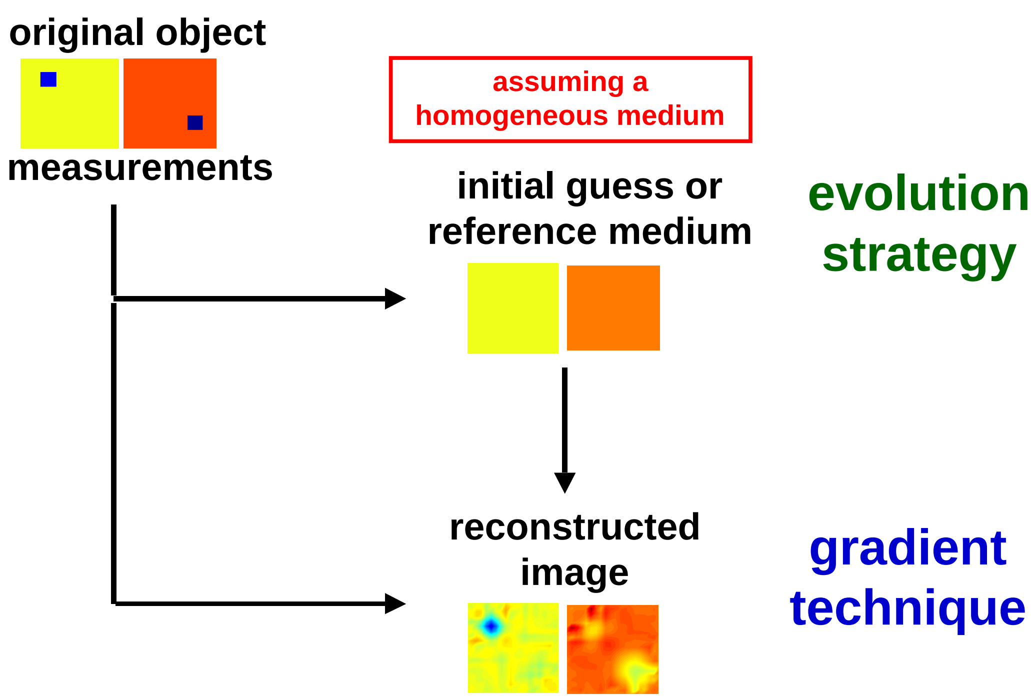

Figure 6. Flow chart of the hybrid approach. The evolution strategy calculates a good initial guess for the gradient technique by assuming a homogeneous medium. Once a homogeneous medium is found that best fits the data of the unknown hererogeneous medium, a gradient-based minimization scheme is employed to reconstruct the heterogeneous structure of the medium.

Downloaded from SPIE Digital Library on 23 Nov 2009 to 128.59.151.11. Terms of Use: http://spiedl.org/terms

3. RESULTS

To test the novel hybrid scheme we performed simulations on a heterogeneous medium that contained variations in the scattering as well as in the absorption coefficient (Fig. 7 - original). The size of the medium was 4 cm x 4 cm. The optical parameters of the background medium were μ_s = 5.0 cm^{-1} , μ_a = 0.8 cm^{-1}, and g = 0. Two heterogeneities were embedded in this background medium. In one of the heterogeneities the absorption coefficient was reduced to μ_a = 0.4 cm^{-1}, and in the other heterogeneity the scattering coefficient was reduced to μ_s = 2.5 cm^{-1}. Measured data were simulated for 4 sources, one centered at each side of the medium, and 60 detectors that surrounded the medium equally spaced. However, measurements were not taken on the same side where the source was positioned. Therefore only 4x45 = 180 source-detector pairs were considered. The numerically generated measurements were corrupted with 30dB noise.

The reconstructions were performed with the time-independent, finite-difference, discrete-ordinate transport code[5] using a 40x40 grid and 16 ordinates. First, the evolution strategy was used to determine μ_a and μ_s values of a homogeneous medium that minimized the objective function in Eq. 1. We started with a population of three parents having the initial parameters (μ_a =0.2 cm^{-1}, μ_s=10.0 cm^{-1}) , (μ_a=0.6 cm^{-1} , μ_s=4 cm^{-1}) , and (μ_a =0.7 cm^{-1}, μ_s=6 cm^{-1}). The anisotropy factor was assumed constant (g = 0). Using the same recombination and mutation rules as in Section 2.3 (Eq. 4 - 6), the parents produced 20 offspring at each generation cycle. The three offspring with the smallest objective function became parents of the next generation. The optimization process stopped after 31 generations, where no further improvement in updating the optical parameters could be found. The final result was μ_a = 0.78 cm^{-1} and μ_s = 5.0 cm^{-1}, which is close to the values of the optical properties of the background medium. The optical parameters determined with the evolution algorithm served as the homogeneous initial guess for the gradient-based reconstruction algorithm. This second step in the reconstruction procedure stopped after 17 iterations, when the convergence criterion ($|\partial\Phi/\partial(\mu_a,\mu_s)|<0.0001$) was reached. Figure 7 shows the result of this hybrid reconstruction. Also shown are two other reconstructions that only used the gradient-based scheme. For these reconstructions two different initial guesses were used. It can be seen that the hybrid approach results in the best reconstruction result. The pure gradient-based scheme only achieves a satisfactory reconstruction when a good initial guess is provided.

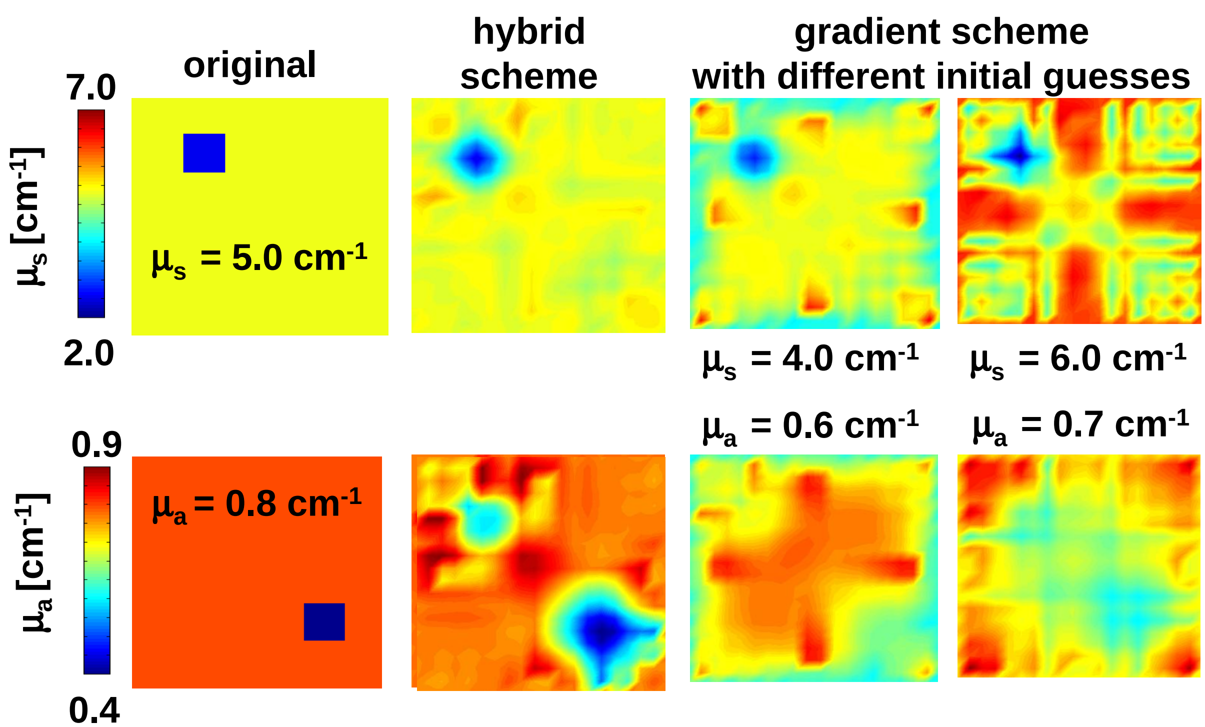

Figure 7. Comparison of gradient-evolution hybrid algorithm and pure gradient-based algorithm. The original spatial distributions of μ_s ,μ_a are shown to the left.

Downloaded from SPIE Digital Library on 23 Nov 2009 to 128.59.151.11. Terms of Use: http://spiedl.org/terms

4. SUMMARY

The image reconstruction problem in diffuse optical tomography (DOT) can be considered as a nonlinear optimization problem. The goal is to find the minimum of an objective function that evaluates the difference between measured and predicted data. The predicted data is calculated with a forward model, which depends on the spatial distribution of optical properties of the medium. Most of the existing image reconstruction algorithms for DOT use gradient-based optimization techniques. Gradient techniques start from an initial guess of the spatial distribution of optical properties and perform iterative line-minimizations along the gradient of the objective function with respect to the optical parameters. The performance of these algorithms depends critically on a good initial guess. If the initial guess is not close to the solution, gradient techniques may not converge or converge to a local minimum.

Evolution strategies are global optimization techniques that mimic the evolutionary process in nature. Using these strategies one applies different operations, such as recombination, mutation, and selection, to a parent population to obtain new generations. The parents are represented by their phenotype, which are in our case given by the optical parameters. Throughout the evolution process the parents create offspring with different optical parameters. The best-adapted offspring (offspring with the smallest objective function) become parents of the next generation. Since no gradient information is used during this process, these schemes are less sensitive to the initial guess and typically can find a global minimum more reliably. However, the computational burden of evolution techniques rapidly increases with the number of unknown parameters in the problem.

In this work we developed a hybrid approach that combines the advantages of low computational burden of a gradient technique with the global-minimum-search capabilities of an evolution strategy. The evolution strategy is used to calculate a good homogeneous initial guess for the gradient technique. Once provided with a good initial guess the gradient-based optimizer is used to determine the location of the heterogeneity. Our initial studies on simulated data show that the evolution-based algorithm finds initial guesses that lead to good convergence of the gradient-based scheme.

ACKNOWLEDGEMENTS

This work was supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, a part of the National Institute of Health (grant # R01 AR46255-01), the Whitaker Foundation (grant # 98-0244), and the Dean of the College of Medicine at the State University of New York - Downstate Medical Center.

REFERENCES

1. S. S. Saquib, K. M. Hanson, and G. S. Cunningham, "Model-based image reconstruction from time-resolved diffusion data," in *Medical Imaging: Image Processing*, Proc. of the SPIE-The International Society for Optical Engineering, vol. 3034, pp. 369-380, 1997.
2. A.H. Hielscher, A. Klose, D.M. Catarious Jr., K.M. Hanson, "Tomographic imaging of biological tissue by time-resolved, model-based, iterative, image reconstruction," OSA Trends in Optics and Photonics: Advances in Optical Imaging and Photon Migration II, Vol. 21, R.R Alfano and J.G. Fujimoto, eds., Optical Society of America, Washington, DC, pp. 156-161, 1998.
3. S.R. Arridge and M. Schweiger, "A gradient-based optimisation scheme for optical tomography," *Optics Express* **2** (6), 213-226 (1998).
4. A.H. Hielscher, A.D. Klose, K.M. Hanson, "Gradient-based iterative image reconstruction scheme for time-resolved optical tomography," *IEEE Transactions on Medical Imaging* **18**, 262-271 (1999).
5. A.D. Klose and A.H. Hielscher, "Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer," *Medical Physics* **26** (8), 1698-1707 (1999).
6. R. Roy and E.M. Sevick-Muraca,"Truncated Newton's optimization scheme for absorption and fluorescence optical tomography: Part I Theory and formulation, *Optics Express* **4** (10), 353-371 (1999).
7. J.C.Ye, K.J. Webb, C.A. Bouman, R.P Millane, "Optical diffusion tomography by iterative-coordinate-descent optimization in a Bayesian framework," *J. Opt. Soc. of Am. A: Optics and Image Science* **16**, 2400-2413 (1999).
8. S.R. Arridge, "Optical tomography in medical imaging," Inverse Problems 15, pp. R41-R93 (1999).

Downloaded from SPIE Digital Library on 23 Nov 2009 to 128.59.151.11. Terms of Use: http://spiedl.org/terms

9. A.H. Hielscher, A.D. Klose, and J. Beuthan, "Evolution-strategies for optical tomographic characterization of homogeneous media", *Optics Express* 7(13), pp. 507-518, 2000. (http://www.opticsexpress.org/framestocv7n13.htm)
10. A.J. Welsh and M.J.C. van Gemert, *Optical-thermal response of laser-irradiated tissue*, Plenum Press, New York, 1995.
11. Stephen G. Nash, *Linear and Nonlinear Programming*, McGraw-Hill
12. H.P. Schwefel, *Evolution and Optimum Seeking*, John Wiley & Sons, New York, NY 1995.
13. T. Bäck, H.P. Schwefel, "An overview of evolutionary algorithms for parameter optimization*," Evolutionary Computation* **1** (1),1-23 1993.
14. Z. Michlewicz, Genetic Algorithms + Data Structures = Evolution Programs, Springer,New York, NY 1999.

Downloaded from SPIE Digital Library on 23 Nov 2009 to 128.59.151.11. Terms of Use: http://spiedl.org/terms