SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Optimization of optode configurations for fluorescence optical tomography by experimental design methods

Manuel Freiberger Christian Clason

Hermann Scharfetter

SFB-Report No. 2009–014 April 2009

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

Der Wissenschaftsfonds.

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Optimization of optode configurations for fluorescence optical tomography by experimental design methods

Manuel Freiberger[a], Christian Clason[b] and Hermann Scharfetter[a]

[a]Graz University of Technology, Institute of Medical Engineering, Kronesgasse 5/II, 8010 Graz, Austria

[b]University of Graz, Institute for Mathematics and Scientific Computing, Heinrichstr. 36/III, 8010 Graz, Austria

ABSTRACT

Fluorescence optical tomography utilizes a set of light sources placed on or above the sample to excite a fluorophore inside it. From the fluorescent light measured on the boundary the distribution of the fluorophore can be reconstructed. The placement of the source and detector optodes determines the quality of the resultant images in terms of resolution and contrast-to-noise ratio, for example.

This paper addresses the optimization of the measurement configuration based on finite element simulations. As quality criterion for optodes, the redundancy of the measurement was chosen. This quantity is easily computed from the Jacobian of the mathematical formulation of light propagation. The implemented algorithm starts with a feasible pool of optodes and searches a subset with minimum redundancy between the measurements in an iterative fashion.

Further author information: (Send correspondence to Manuel Freiberger)

Manuel Freiberger: E-mail: manuel.freiberger@tugraz.at

Christian Clason: E-mail: christian.clason@uni-graz.at

Hermann Scharfetter: E-mail: hermann.scharfetter@tugraz.at

The algorithm allows to bias the design in order to favor reconstruction results in an *a-priori* given region.

Two different variations of the algorithm were implemented: one based on geometric averaging and on arithmetic averaging. Both deliver similar optimal optode configurations. The arithmetic averaging shows a slightly better condition number of the sensitivity matrix while the geometric averaging approach corresponds better with entropy optimization from a mathematical point of view.

We present optimal illumination and detector patterns for an initial set of 48 source and 48 detector optodes placed in a regular grid on a cylinder with focusing on different regions. Further, examples for the attenuation of fluorophore signals from regions outside the focus are given.

Keywords: Fluorescence optical tomography, hardware optimization, redundancy minimization, entropy optimization

1. INTRODUCTION

Fluorescence diffusion optical tomography (FDOT) is one of the newer imaging techniques with promising application potential in medicine. FDOT provides the possibility of functional imaging, *i.e.* it not only visualizes anatomical structures but also provides information about physiological states and processes.

FDOT utilizes the ability of fluorescent dyes to absorb light in a certain wavelength range and to emit photons at a higher wavelength. The excitation light is injected into the sample through a set of sources. A source can either be in contact with the sample's surface (*e.g.* a waveguide) or it delivers the light in a contactless manner using collimated or divergent lightbeams. The excitation light is scattered and absorbed while spreads in the tissue. At sites where a fluorophore is present and active (*e.g.* inside a tumor), a part of the absorbed light leads to

re-emission at another wavelength. This secondary light is again scattered through the tissue and the part which reaches the boundary can be measured by photo detectors.

Due to the diffuse propagation of the photons in tissue,[1] light emerging from the fluorescent dye widely spreads before it reaches the boundary. This is in contrast to other established imaging techniques like X-rays where the rays travel through the sample of interest in nearly straight lines. The photon diffusion has to be considered in a suitable forward model which is the basis for the reconstruction algorithm that seeks to determine the distribution of the fluorophore from boundary measurements.

The reconstructed results usually improve when increasing the number of sensors. However, this is only true up to a certain extent as the diffuse nature of the photon propagation inherently limits the independence of information of different sensors and hence the obtainable resolution. Contrary the cost for the detector hardware, the acquisition time and the computation effort as well as the memory needed for reconstruction increase as the number of source/detector combinations grow larger. The goal is to find a good compromise between image quality and both hardware and reconstruction feasibility.

Graves *et al*[2] have performed some investigations on how the number of sources and detectors and their distance, respectively, influences the reconstruction. Later on the method was extended by Lasser *et al*,[3] who applied it to 360° projection tomography.

This paper presents a different approach to optimize the optode configuration of a fluorescence tomography system in the sense that it does not just compare different optode configurations but provides an information measure for every single optode offering greater flexibility, therefore. Furthermore, it will be shown how the optimization can be modified such that the reconstruction is focused on a given region of interest.

2. METHODS

2.1 Forward model

One of the most accurate ways to model light propagation is to utilize Boltzmann's transport equation for kinetic gases. The photons can then be treated like independent gas particles leading to the radiation transfer equation. Unfortunately, the photon intensity in the radiation transfer equation is a field dependent on the spatial coordinates and the direction (*i.e.* two angles) into which the photons travel. This leads to a discretization with a huge number of degrees of freedom and requires extensive computing power and memory.

Therefore, it is common to use an approximation of the transfer equation known as diffusion equation.[4] Including a spatially variable fluorophore concentration c the diffusion equation reads

$$-\nabla \cdot (\kappa(x,\lambda)\nabla\varphi(x)) + \left(\mu_{a,i}(x,\lambda) + c(x)\varepsilon(\lambda) + \frac{i\omega}{v}\right)\varphi(x) = q(x), \tag{1}$$

together with the boundary condition

$$\varphi(x) + 2R\kappa(x,\lambda)\frac{\partial\varphi(x)}{\partial n} = 0 \tag{2}$$

where φ is the photon density field, $\kappa = (3(\mu_{a,i}+c\varepsilon+\mu_s'))^{-1}$ is the diffusion coefficient, μ_a and μ_s' are the intrinsic absorption and reduced scattering coefficient, respectively, ε is the fluorophore's extinction which relates the fluorophore concentration to photon absorption. λ is the wavelength of the light, ω is the modulation frequency of the light source, v is the speed of light and q is the source term. R is a factor to incorporate reflections at the boundary whose normal is denoted by n. Equation (1) can be solved efficiently using a finite element discretization.

For fluorescence applications, two diffusion equations—one describing the propagation of the excitation photons ($\lambda = \lambda_{ex}$, $\varphi = \varphi_{ex}$), another one to describe the emission field ($\lambda = \lambda_{em}$, $\varphi = \varphi_{em}$)—can be coupled. We prefer to write this in an operator (or matrix-like) notation,

where A_{ex} and A_{em} describe the propagation of the excitation and emission field, respectively.

$$A_{ex}(c)\varphi_{ex} = q$$
$$A_{em}(c)\varphi_{em} = B(c)\varphi_{ex} \tag{3}$$

The operator B converts the photon density from the excitation wavelength to the emission wavelength at sites where a fluorophore is present and thus serves as a source for the emission field. It is defined as

$$B(c)\varphi_{ex}(x) = \frac{Q}{1 - i\omega\tau} c(x)\varepsilon_{ex}\varphi_{ex}(x), \tag{4}$$

whereby Q is the quantum yield and τ the fluorescence life-time.[5]

Although more elaborate detector models (*e.g.* reference 6) could be used, in this paper a measurement d is defined as the number of photons leaving the sample at a certain point x_D per unit time:

$$d := -v \int_{\partial\Omega} \delta(x - x_D)\kappa_{em}(x)\frac{\partial\varphi_{em}(x)}{\partial n} dx \overset{(2)}{=} v \int_{\partial\Omega} \frac{1}{2R}\delta(x - x_D)\varphi_{em}(x) dx. \tag{5}$$

If more than one source and one detector are present, we denote the measurement made with the i-th pair of source/detector by d_i.

2.2 Sensitivity

In order to solve the inverse problem, *i.e.* the reconstruction of the distribution of the fluorophore's concentration $c(x)$ from measurements on the boundary, it is necessary to know the influence of a change in the concentration distribution on the measurements. In other words the so-called *sensitivity*, the derivative of the system (3) with respect to $c(x)$, is needed. Since the measurement d is a linear functional of the emitted photon density φ_{em}, it suffices to calculate the derivative of φ_{em} with respect to c. To this end, we write the coupled system of partial differential equations describing the dependence of φ_{em} (and φ_{ex}) on c as a nonlinear operator

equation $F(c, \varphi(c)) = 0$, where $\varphi = (\varphi_{ex}, \varphi_{em})$ and

$$F : (c, \varphi) \to \begin{cases} A_{ex}(c)\varphi_{ex} - q, \\ A_{em}(c)\varphi_{em} - B(c)\varphi_{ex}. \end{cases} \tag{6}$$

Since F is continuous and Fréchet-differentiable with respect to c and φ, the Banach space version of the implicit function theorem[7] states that $\varphi(c)$ is Fréchet-differentiable with respect to c, and that the derivative $\varphi'(c)$ satisfies

$$\partial_\varphi F(c, \varphi(c))\, \varphi'(c) = -\partial_c F(c, \varphi(c)) \tag{7}$$

provided that $\partial_\varphi F(c, \varphi(c))$ is bijective. The operator F being linear in φ, the partial Fréchet derivative with respect to φ applied to the variation $\delta\varphi = (\delta\varphi_{ex}, \delta\varphi_{em})$ is just

$$\partial_\varphi F(c, \varphi(c))\, \delta\varphi = \begin{cases} A_{ex}(c)\delta\varphi_{ex}, \\ A_{em}(c)\delta\varphi_{em} - B(c)\delta\varphi_{ex}, \end{cases} \tag{8}$$

and the bijectivity follows from the well-posedness of the forward problem.

It remains to calculate the Fréchet derivative $\partial_c F(c, \varphi(c))$ acting on the variation δc. Taking the derivative of the system (3) with respect to c and setting

$$\kappa'_{ex} = \frac{\varepsilon_{ex}}{3(\mu_{a,i,ex} + c\varepsilon_{ex} + \mu'_s)^2}, \tag{9}$$

and correspondingly for κ'_{em}, we find that

$$\partial_c F(c, \varphi(c))\, \delta c = \begin{cases} -\nabla \cdot (-\kappa'_{ex}\delta c \nabla\varphi_{ex}) + \delta c \varepsilon_{ex}\varphi_{ex}, \\ -\kappa'_{ex}\delta c \partial_n \varphi_{ex}, \\ -\nabla \cdot (-\kappa'_{em}\delta c \nabla\varphi_{em}) + \delta c \varepsilon_{em}\varphi_{em} - \dfrac{Q}{1 - i\omega\tau}\delta c \varepsilon_{ex}\varphi_{ex}, \\ -\kappa'_{em}\delta c \partial_n \varphi_{em}. \end{cases} \tag{10}$$

Therefore, in order to calculate the sensitivity of the measurement d for given c with respect to a perturbation δc, we first compute $\varphi_{ex}(c)$ and $\varphi_{em}(c)$ as the solution of (3) and then solve

the boundary value problem

$$\begin{cases} -\nabla\cdot(\kappa_{ex}\nabla\delta\varphi_{ex}) + (\mu_{a,i,ex} + c\varepsilon_{ex} + \mu_s')\,\delta\varphi_{ex} = -\nabla\cdot(\kappa_{ex}'\delta c\nabla\varphi_{ex}) - \delta c\varepsilon_{ex}\varphi_{ex}, \\ \varphi_{ex} + 2R\kappa_{ex}\partial_n\varphi_{ex} = 2R\kappa_{ex}'\delta c\partial_n\varphi_{ex} \end{cases} \tag{11}$$

for $\delta\varphi_{ex}$, followed by

$$\begin{cases} -\nabla\cdot(\kappa_{em}\nabla\delta\varphi_{em}) + (\mu_{a,i,em} + c\varepsilon_{em} + \mu_s')\,\delta\varphi_{em} = \\ \qquad -\nabla\cdot(\kappa_{em}'\delta c\nabla\varphi_{em}) - \delta c\varepsilon_{em}\varphi_{em} + \dfrac{Q\varepsilon_{ex}}{1-i\omega\tau}\left(\delta c\varphi_{ex} + c\delta\varphi_{ex}\right), \\ \varphi_{em} + 2R\kappa_{em}\partial_n\varphi_{em} = 2R\kappa_{em}'\delta c\partial_n\varphi_{em} \end{cases} \tag{12}$$

for $\delta\varphi_{em}$. The change in measurement at a certain location x_D is then given by

$$\delta d = v\int_{\partial\Omega}\frac{1}{2R}\delta(x-x_D)\delta\varphi_{em}(x)dx. \tag{13}$$

In a finite element context, the discretization of the concentration using piecewise-constant ansatz functions in (3) and (11)–(12) leads to the Jacobian or sensitivity matrix which is denoted by J in this paper. The element J_{ij} describes the effect of a concentration change in the j-th finite element on the i-th measurement.

In certain applications, it is feasible to operate with difference measurements. A measurement d_0 is made with a base-line concentration c_0 and a second measurement d_1 is performed after the concentration distribution has changed to c_1. If the difference in concentration is small, the following linearization can be used for reconstruction:

$$\Delta d = J(c_0)\Delta c, \tag{14}$$

where Δd is a vector of difference measurements and Δc the vector of concentrations in the finite elements. This formulation will be used throughout this paper. However, the Δ is neglected from now on and we understand all measurement and concentrations as differences from a base-state.

3. OPTIMIZATION OF THE MEASUREMENT SETUP

Entropy-based optimization methods have a quite long history in image processing and reconstruction[8] and have been applied to various fields of tomography as can be seen from references 9, 10, 11, 12, 13, just to name a few. The basic idea is to treat the unknown parameter—in our case the fluorophore concentration—as random variable with a certain probability density. Then one seeks to reconstruct that parameter distribution leading to the maximum entropy, for example.

The optimization approach followed in this paper is based on the idea that the different measurements should be as independent as possible, *i.e.* every measurement should result in new information which can be used for the inverse problem. A way to quantify this independence is by using the mutual information (MI).

Let $\mathcal{M}$ denote the set of all measurement indices, *i.e.* each element of $\mathcal{M}$ uniquely defines one pair of source/detector. Further, let $\mathcal{S}_i \subset \mathcal{M}$ be the indices of those measurements which are made with the i-th source. Without loss of generality it can be assumed that the measurements are ordered such that for one fixed source i the sensitivity matrix and the measurements can be partitioned as

$$J = \begin{pmatrix} J_1 \\ J_2 \end{pmatrix}, \quad \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} J_1 \\ J_2 \end{pmatrix} c \tag{15}$$

where d_2 consists of all measurements made with the i-th source and d_1 those made with other sources. The sub-matrix J_1 has got $|\mathcal{M} \setminus \mathcal{S}_i|$ rows and J_2 has got $|\mathcal{S}_i|$ giving a total number of $|\mathcal{M}|$ for J.

In order to calculate the entropy, one has to interpret the model parameters as random variables and one has to make assumptions on their probability distribution. For the sake of simplicity, it is assumed that the model parameters (*i.e.* the concentrations in the finite elements) are independent and normally distributed with equal variance σ_c. This will render the model covariance matrix diagonal: $\mathrm{cov}(c) = \sigma_c I$. The data covariance matrix is coupled via the relation

$d = Jc$ and it applies that

$$\operatorname{cov}(d) = \operatorname{cov}(Jc) = J \operatorname{cov}(c) J^T = \sigma_c J J^T. \tag{16}$$

or using the split data set:

$$\operatorname{cov} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \sigma_c \begin{pmatrix} J_1 J_1^T & J_1 J_2^T \\ J_2 J_1^T & J_2 J_2^T \end{pmatrix}. \tag{17}$$

The entropy (or uncertainty) of the full data is given through the multivariate normal distribution[14] as

$$H(d) = \frac{1}{2} \log \left[(2\pi e)^M \det(\operatorname{cov}(d)) \right] \tag{18}$$

For the split data set, the entropy if only d_1 is measured is

$$H(d_1) = \frac{1}{2} \log \left[(2\pi e)^{M-S_i} \det(\operatorname{cov}(d_1)) \right] \tag{19}$$

and the conditional entropy of d_1 if d_2 is already known is

$$H(d_1|d_2) = \frac{1}{2} \log \left[(2\pi e)^{M-S_i} \det(\operatorname{cov}(d_1|d_2)) \right], \tag{20}$$

where $\operatorname{cov}(d_1|d_2)$ is the conditional covariance. It can be calculated from equation (17) by the Schur complement of $J_2 J_2^T$ in $J J^T$:[15]

$$\operatorname{cov}(d_1|d_2) = \sigma_c \left(J_1 J_1^T - J_1 J_2^T (J_2 J_2^T)^{-1} J_2 J_1^T \right) \tag{21}$$

This is well defined, because $J_2 J_2^T$ is invertible due to the fact that J has a lot more columns than rows and there are no two measurements with the same sensitivity distribution which means that the rows of J—and those of J_1 and J_2, thus—are linearly independent.

As a remark we note that the term $J_2^\dagger := J_2^T (J_2 J_2^T)^{-1}$ appearing in equation (21) is the pseudo-inverse of J_2. If $J_2^\dagger J_2 = I$ would be fulfilled exactly, the covariance of d_1 would be zero. In such a case the model parameters would lead to measurement data d_2 that contains already all the information and d_1 can be predicted from it. Furthermore, all model parameters could be reconstructed exactly from the knowledge of the measurements d_2 alone.

Now, a source can be removed safely, if the conditional entropy $H(d_1|d_2)$ is low because in that case measuring d_2 decreases the uncertainty in d_1 a lot. In other words, the information in the measurements d_1 (made with the source optode under test) is also widely explained by the measurements d_2 (made without this source). This can also be expressed by introducing the mutual information $\mathrm{MI}(d_1, d_2) := H(d_1) - H(d_1|d_2)$ which quantifies the reduction of uncertainty in d_1 when d_2 is known beforehand. If the measurements d_2 of the currently considered source have a high mutual information with all other measurements, the source can be removed from the pool as its information is also present in d_1 to a large extent.

Writing the negative mutual information together with the equations (19)–(21) gives

$$-\mathrm{MI}(d_1, d_2) = \frac{1}{2} \log \left[\det \left(J_1 J_1^T - J_1 J_2^T (J_2 J_2^T)^{-1} J_2 J_1^T \right) / \det(J_1 J_1^T) \right] \tag{22}$$

Using the same argumentation as for $J_2 J_2^T$, $J_1 J_1^T$ is also invertible. Furthermore it is required that the matrix has an invertible square root matrix such that

$$\det{}^{-1}(J_1 J_1^T) = \det \left((J_1 J_1^T)^{-\frac{1}{2}} \right) \det \left((J_1 J_1^T)^{-\frac{1}{2}} \right). \tag{23}$$

Eventually, we arrive at the expression

$$-\mathrm{MI}(d_1, d_2) = \frac{1}{2} \log \left[\det \left(I - (J_1 J_1^T)^{-\frac{1}{2}} J_1 J_2^T (J_2 J_2^T)^{-1} J_2 J_1^T (J_1 J_1^T)^{-\frac{1}{2}} \right) \right] \tag{24}$$

A major drawback of this technique is the fact that the computation of the mutual information requires quite some computing effort due to the necessity of matrix inversions and the computations of the determinants in equation (24). This renders such an approach computationally infeasible, which is why an alternative method was implemented instead.

3.1 Redundancy reduction

A different method was originally developed by Michelini and Lomax and published by Curtis *et al.*[16] They quantified the independence of two measurements by computing the inner product

and the angle, respectively, between the respective rows of the sensitivity matrix J. Then the algorithm has to find that set of measurements which is closest to an orthogonal set.

Using the same notation as in the previous section, the square of the cosine of the angle between two measurements, one made with source i and one made with another source, is given by the term

$$\frac{(j_m, j_n)^2}{\|j_m\|^2 \, \|j_n\|^2}, \quad m \in \mathcal{S}_i, \quad n \in \mathcal{M} \setminus \mathcal{S}_i, \tag{25}$$

where j_m and j_n denote the m-th and n-th row of the sensitivity matrix J, respectively. This quantity is a real number from the interval $[0, 1]$. If the measurements are orthogonal, the cosine will be zero and the more they depend on each other, the more the cosine will approach 1. In contrast to the the formula in 16 the square of the cosine is used here which turns out to be advantageous for the comparison to the mutual information measure later on.

Now, consider the average square of the cosine between all measurements made with source i and those measurements made with another source. This will lead to the expression

$$r_i := \frac{1}{|\mathcal{M} \setminus \mathcal{S}_i|} \sum_{n \in \mathcal{M} \setminus \mathcal{S}_i} \frac{1}{|\mathcal{S}_i|} \sum_{m \in \mathcal{S}_i} \frac{(j_m, j_n)^2}{\|j_m\|^2 \, \|j_n\|^2} = \frac{1}{|\mathcal{M} \setminus \mathcal{S}_i| \, |\mathcal{S}_i|} \sum_{n \in \mathcal{M} \setminus \mathcal{S}_i} \sum_{m \in \mathcal{S}_i} \frac{(j_m, j_n)^2}{\|j_m\|^2 \, \|j_n\|^2}, \tag{26}$$

which is again from $[0, 1]$. If the value is close to 1, the information gained using source i can also be gained using other sources, which means that the source i is redundant. Therefore, we call r_i the *redundancy* of source i. The quantity

$$q_i := 1 - r_i = 1 - \frac{1}{|\mathcal{M} \setminus \mathcal{S}_i| \, |\mathcal{S}_i|} \sum_{n \in \mathcal{M} \setminus \mathcal{S}_i} \sum_{m \in \mathcal{S}_i} \frac{(j_m, j_n)^2}{\|j_m\|^2 \, \|j_n\|^2} \tag{27}$$

is used as a quality criterion in the optimization algorithm. An analog expression can be derived for the detectors. This is achieved by replacing $\mathcal{S}_i$ in (26) by $\mathcal{D}_i$, the set of indices belonging to the i-th detector.

The optimization algorithm starts with a set of feasible optodes. It then calculates the quality measure for every source and removes the one with the lowest measure from the optode pool. In the next iteration the same is performed for the detector optodes. This is done until a

given stopping criterion is met. The optodes left in the pool are considered to be the optimal measurement configuration.

3.1.1 Geometric averaging

The averaging of the single optode redundancies in the former section were introduced in an intuitive way using the arithmetic mean. However, one could also think about using the other methods like the geometric mean, for example:

$$q_{g,i} := \left[\prod_{n \in \mathcal{M} \setminus \mathcal{S}_i} \left(1 - \frac{1}{|\mathcal{S}_i|} \sum_{m \in \mathcal{S}_i} \frac{(j_m, j_n)^2}{\|j_m\|^2 \, \|j_n\|^2} \right) \right]^{\frac{1}{|\mathcal{M} \setminus \mathcal{S}_i|}} \tag{28}$$

The advantage of this formulation is that from a mathematical point of view it is linked more closely to entropy optimization as will be shown in section 3.2.

3.2 Relation to entropy optimization

In this section a link between the mutual information optimization and the redundancy minimization technique is established. To relate the redundancy to entropy, equation (26) has to be brought into a matrix formulation. Using the partitioned sensitivity defined in equation (15) and the notation

$$\operatorname{diag}^{-\frac{1}{2}} \left(J J^T \right) = \begin{pmatrix} \frac{1}{\|j_1\|} & & \\ & \ddots & \\ & & \frac{1}{\|j_M\|} \end{pmatrix}, \tag{29}$$

the redundancy from equation (26) can be re-written as

$$r_i = \frac{1}{|\mathcal{M} \setminus \mathcal{S}_i| \, |\mathcal{S}_i|} \sum_{k=1}^{|\mathcal{M} \setminus \mathcal{S}_i|} \sum_{l=1}^{|\mathcal{S}_i|} \left(\operatorname{diag}^{-\frac{1}{2}} \left(J_1 J_1^T \right) J_1 J_2^T \operatorname{diag}^{-\frac{1}{2}} \left(J_2 J_2^T \right) \right)_{k,l}^2 . \tag{30}$$

This is actually the square of the Frobenius norm of the matrix $\operatorname{diag}^{-\frac{1}{2}} \left(J_1 J_1^T \right) J_1 J_2^T \operatorname{diag}^{-\frac{1}{2}} \left(J_2 J_2^T \right)$ and can also be expressed by the trace of the matrix multiplied with its transposed. Thus, the matrix form of equation (26) reads:

$$r_i = \frac{1}{|\mathcal{M} \setminus \mathcal{S}_i| \, |\mathcal{S}_i|} \operatorname{tr} \left(\operatorname{diag}^{-\frac{1}{2}} \left(J_1 J_1^T \right) J_1 J_2^T \operatorname{diag}^{-1} (J_2 J_2^T) J_2 J_1^T \operatorname{diag}^{-\frac{1}{2}} \left(J_1 J_1^T \right) \right). \tag{31}$$

Finally the matrix form of the quality measure from equation (27) is:

$$q_i = \frac{1}{|\mathcal{M} \setminus \mathcal{S}_i|} \operatorname{tr}\left(I - \frac{1}{|\mathcal{S}_i|} \operatorname{diag}^{-\frac{1}{2}}\left(J_1 J_1^T\right) J_1 J_2^T \operatorname{diag}^{-1}(J_2 J_2^T) J_2 J_1^T \operatorname{diag}^{-\frac{1}{2}}\left(J_1 J_1^T\right)\right). \tag{32}$$

A similar derivation can be done for the geometric quality measure introduced in equation (28). The final result will be:

$$q_{g,i} = \left[\det \operatorname{diag}\left(I - \frac{1}{|\mathcal{S}_i|} \operatorname{diag}^{-\frac{1}{2}}\left(J_1 J_1^T\right) J_1 J_2^T \operatorname{diag}^{-1}(J_2 J_2^T) J_2 J_1^T \operatorname{diag}^{-\frac{1}{2}}\left(J_1 J_1^T\right)\right)\right]^{\frac{1}{|\mathcal{M}\setminus\mathcal{S}_i|}}. \tag{33}$$

When the results of equations (32), (33) and (24) are compared, one notices interesting similarities in the structure of these equations although they are not the same. The mutual information formula (24) combines the whole conditional covariance matrix into a single quality measure using the determinant. The other two equations based on the redundancy operate on the diagonal matrix parts (the variances) only and neglect the covariances completely.

The original formulation based on the arithmetic mean (equation (32)) has the disadvantage that there is no strong relationship between the trace of a (symmetric positive-definite) covariance matrix and its determinant in general. In fact it is rather easy to construct examples where the trace between two setups increases while the determinant decreases.

The newly introduced geometric averaging (equation (33)) resembles the entropy optimization much more closely. By using the Cauchy-Schwarz inequality $\operatorname{cov}^2(x, y) \leq \operatorname{var}(x) \operatorname{var}(y)$, it is obvious that a reduction in the variances have to reduce the covariances simultaneously. Thus, we can argue that a decrease in the geometrically weighted redundancy quality measure $q_{g,i}$ will decrease ($-$ MI) and therefore increase the mutual information. In other words, an optode which is highly redundant is likely to exhibit a high mutual information content between measurement associated to that optode and all other measurements.

3.3 Focusing

In certain applications it can be advantageous to bias the arrangement of the optodes in order to reach a higher sensitivity—which usually goes along with a higher resolution and/or a better

Table 1. Values of optical parameters used for the forward simulation[5,17,18]

	μ_s'	$\mu_{a,i}$	$c \cdot \varepsilon$	R
	mm^{-1}	mm^{-1}	mm^{-1}	
excitation	0.275	0.036	$0.84 \cdot 10^{-3}$	2.51
emission	0.235	0.029	$0.25 \cdot 10^{-3}$	2.51

contrast-to-noise ratio—in a specified region. This can be used to focus the reconstruction on certain organs, for example.

A simple approach to achieve focusing is by weighting the columns of the sensitivity matrix with a pre-defined weighting mask f:

$$J_F = J\,\mathrm{diag}(f), \tag{34}$$

whereby J_F denotes the focused sensitivity matrix. This resultant matrix will then be used in the optimization algorithm described above.

In the simplest case f is a binary vector being one in the region of interest and zero everywhere else. Also smooth variations of the mask are possible. Generally we can assume that $0 \leq f_i \leq 1$.

4. RESULTS

The optode optimization was performed on a cylinder with a height of 90 mm and a radius of 30 mm, which should mimic a small animal. The values of the optical properties can be found in table 1. The fluorophore concentration was assumed to be spatially constant and rather low. This should model a cylinder without fluorescent dye but with a certain amount of auto-fluorescence. Equations and estimates of these parameters can be found in references 5, 17, 18. The absorption due to auto-fluorescence was chosen to be 10^{-2} of the fluorophore absorption found in the thesis by Joshi.[5]

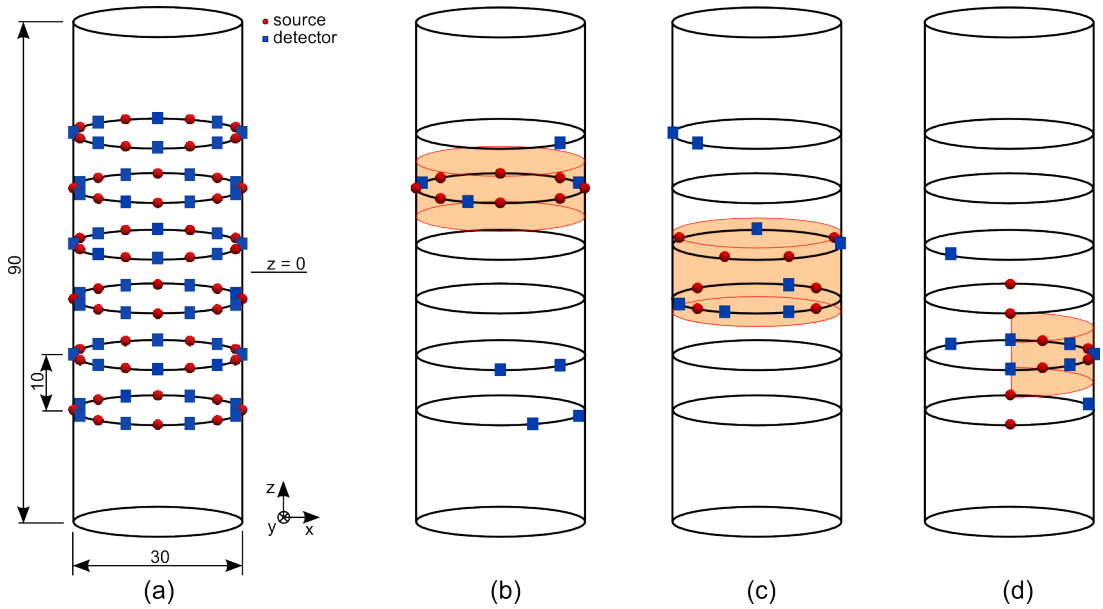

Figure 1. (a) Geometry of the optimization model (measures in mm) together with the initial pool of feasible optodes which are arranged in a zig-zag pattern on six rings with 10 mm spacing. (b)–(d) The result of the optode optimization using the geometric averaging approach for three different focus regions which are drawn in orange.

As initial pool of feasible optodes position a regular grid with 48 source and 48 detector nodes was specified. The optodes were arranged in a zig-zag like pattern on six rings with a spacing of 10 mm (see figure 1(a)). The optimization algorithm needs the desired number of sources and detectors as stopping-criterion. It was chosen to search for an optimal configuration consisting of eight sources and eight detectors.

Three different focus regions were chosen to demonstrate the optimization. Region A consists of the voxels in the cylinder slice given by $10\,\text{mm} < z < 20\,\text{mm}$, region B is another slice defined by $-7.5\,\text{mm} < z < 7.5\,\text{mm}$ and region C is the half-cylinder slice $-20\,\text{mm} < z < -10\,\text{mm}$ and $x > 0\,\text{mm}$. The optimization was performed on a finite element mesh with approximately 30,000 elements.

As can be seen in the plots 1(b)–1(d), the optimization algorithm tends to concentrate the optodes near the focused region. The source optodes are very close to the region of interest and

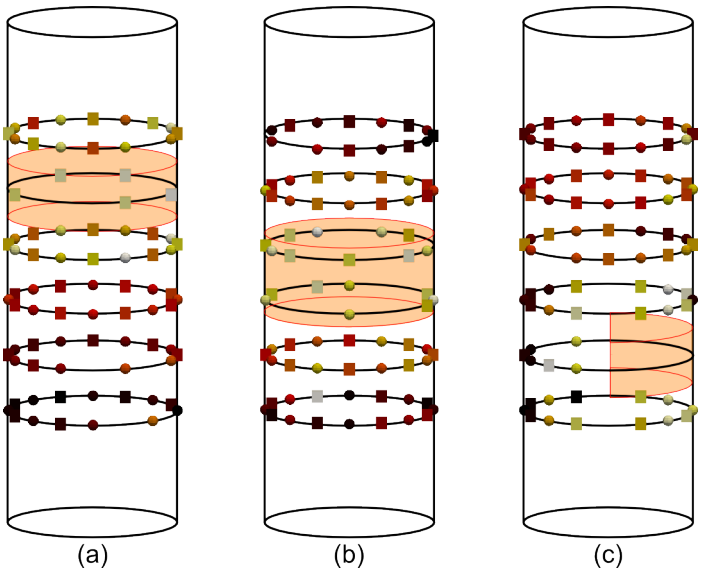

Figure 2. The rank of the removed optodes using the geometrical weighting encoded in color for three different focus regions (sketched in orange). Optodes drawn in black or red were eliminated early during the optimization while those in yellow and white remained longer in the pool. The optimal optodes have been omitted.

so are the detectors in 1(d). However, the detector optodes when focusing on larger slices as in 1(b) and 1(c) are more spread.

Figures 2(a)–2(c) shows the rank of the optodes, *i.e.* the iteration in which they were removed from the pool of feasible optodes. One can see clearly that the optodes far away from the focus region, which is drawn in orange, are removed first while those near or in the region remain longer in the feasible set which is a reasonable result.

We also had the chance to compare the geometric averaging method to the original formulation published in 16. The result of the optimization using the arithmetic averaging is shown in figure 3. As can be seen the distribution of the optodes is nearly the same to a large extent. As the geometric averaging only replaces one of the sums by a product it was not expected to see dramatic differences in the optimal configuration.

As a objective quality measure for the the resultant sensitivity matrix, its singular values and its condition number, respectively, can be used. The largest and smallest singular values

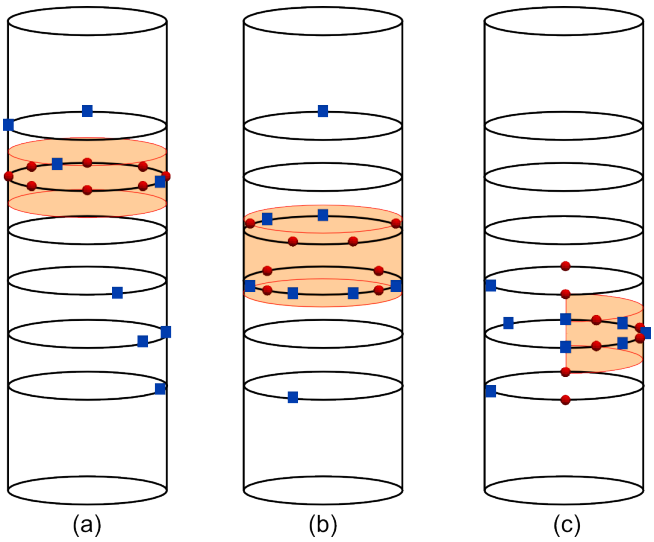

Figure 3. The optimal optode setup for three different focus regions using the geometrically weighted quality measure. The focus regions are sketched in orange.

together with the ratio between them can be found in 2. The full optode configuration has a rather high ratio of $6 \cdot 10^{12}$ and is rather ill-conditioned, thus. The focused designs show a SV ratio which is reduced by a factor of at least 10^6. This is exactly what is intended by the redundancy minimization algorithm as the removal of non-orthogonal rows from the sensitivity matrix improves conditioning. The ratio of the design focusing on region A (the uppermost cylinder slice) is higher by a factor of approximately 100 than the other two designs. This is due to the rather large spread of the optimal detectors (figure 1(b)).

When comparing the geometric to the arithmetic averaging, one notices that the latter exhibits a condition number that is reduced by a factor of 4 approximately meaning that its resultant sensitivity matrix will be even more easy to invert.

Figure 4 shows example reconstructions based on artificial measurement data using the three different focusing strategies together with geometric averaging. The synthetic data was generated using a more fine-grained finite element mesh with three spherical perturbations having a diameter of 5 mm. The optical properties are the same as listed in table 1. For the exact location of the perturbations please refer to figure 4. In figures 4(b) and 4(c) it can bee

Table 2. Singular value (SV) analysis for the full sensitivity matrix and the optimized configurations with focusing to different regions. The table lists the largest and smallest singular value as well as the ratio between them (the condition number) which is a measure of stability for matrix inversion. Geometric averaging refers to the formulation introduced in this paper (see equation(33)) while arithmetic averaging is the original formulation by Curtis *et al* (equation(32)).

Averaging	Design	max SV	min SV	ratio
	Full pool of optodes	60.21e-3	9.972e-15	6.066e12
Geometric	Focus on region A	27.7491e-3	8.5059e-9	3.2623e+6
	Focus on region B	27.0283e-3	485.3229e-9	55.6914e+3
	Focus on region C	29.5896e-3	2.8534e-6	10.3701e+3
Arithmetic	Focus on region A	27.3989e-3	32.9629e-9	831.2028e+3
	Focus on region B	27.0694e-3	1.6307e-6	16.6003e+3
	Focus on region C	29.4153e-3	3.9581e-6	7.4317e+3

seen that the reconstructed perturbation is shifted with respect to the true position because the signal from the unfocused perturbations is quite strong. This shift is much more reduced when the focus is put on the lower half-cylinder slice (region C) as can be depicted from figure 4(d).

5. DISCUSSION

The location of the sensors and detectors is a critical design parameter for FDOT hardware. A configuration determines the sensitivity of the measurements in a given region and also the obtainable resolution. The method presented herein is based on a simulation model and can be used prior to building hardware, thus.

Comparisons of different hardware configurations for fluorescence tomography that we are aware of were previously reported by Graves *et al*[2] and Lasser *et al*.[3] Their approaches were based on a singular value (SV) analysis of the so-called weight matrix. This matrix is essen-

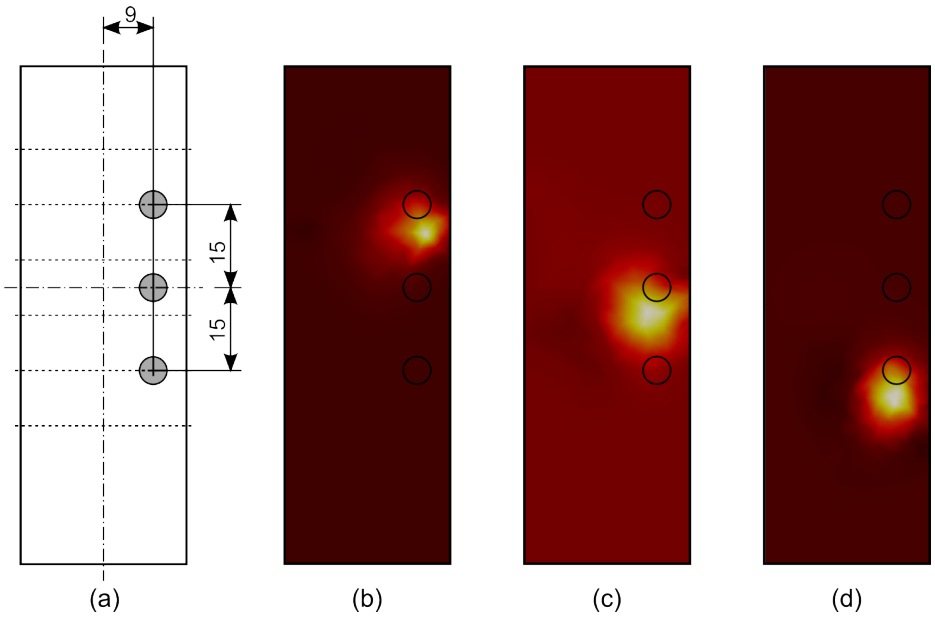

Figure 4. Reconstruction examples for differently focused optode setups. (a) The simulation model used to generate artificial measurement data for all optode setups. The spherical perturbations have a diameter of 5 mm. The six dotted horizontal lines mark the location of the optode rings. The other figures show reconstructions when focusing on region A (b), region B (c) or region C (d) with the optimal setup gained with the geometric averaging method.

tially the sensitivity matrix used in this paper but was obtained using the normalized Born approximation.[19]

In contrast to the previous methods, the optimization algorithm used herein operates on the complete derivative of the diffusion approximation (7). Therefore, also the change of the excitation field due to a perturbation in the fluorophore distribution is considered which is neglected in commonly used first-order Born approximations.

The SV analysis implemented by Graves and Lasser requires a singular value decomposition (SVD) of the sensitivity matrix. This demands a large amount of computation time. As an example, the calculation of a single SVD for a matrix of size $48 \cdot 48 \times 30,000$ takes about 50 minutes on an Intel Core2Duo with two processors. Therefore, the SV optimization is limited to 2D applications or to rather simple 3D geometries, which can be modeled with fewer finite

elements. The redundancy reduction algorithm does not suffer from these limitations as it only needs the sensitivity matrix on which it performs inner products basically. The matrix is built up efficiently using an implicit function formulation which requires one additional finite element solution per optode only. Furthermore, both the assembly of the sensitivity matrix as well as the calculation of the inner products can be parallelized with moderate effort (the basic principle can be found in reference 20, for example) which allows to speed-up the method even more.

The algorithm we implemented starts with a complete optode arrangement and iteratively discards the one optode having the lowest quality measure. Unfortunately, the algorithm cannot determine in the current step whether it would be advantageous later on if the worst optode was kept and another one was dismissed from the pool instead. Thus, the drawback of this simple top-down strategy is that it cannot guarantee to find the global minimum. However, a full search is a NP-problem and would require the computation of around 10^{18} different configurations for the rather simple geometry demonstrated above which is computationally not feasible. A viable alternative could be the implementation of stochastic methods that have a long tradition in numerical optimization.

A great advantage of the redundancy minimization is that it provides a quality measure for every single optode rather than comparing complete configurations as it is the case in SV analysis. This offers the possibility to choose a superior configuration first, which could even be obtained with a completely different method, and to optimize the arrangement further through the removal of optodes exhibiting a poor quality measure.

The large spread of the detector optodes as it is visible especially in figure 1(b) may appear counter-intuitive at first sight. However, still unpublished observations show that the intrinsic absorption $\mu_{a,i}$ influences the optimal configuration. If this parameter is increased, the optimal optodes cluster more around the focused region.

To our knowledge, the possibility of choosing an optode configuration such that the reconstruction is focused to an *a-priori* given region inside the tissue has not been investigated for

fluorescence tomography before. Similar strategies have been reported for other diffusion limited imaging modalities as seismic tomography[16] and magnetic induction tomography,[21] for example.

The effect of focusing on the reconstruction is demonstrated in figure 4. Unfortunately, the first two focusing strategies (figures 4(b) and 4(c)) show a pronounced shift between the true and the reconstructed perturbations. This behavior is significantly reduced when the focus is only put on the half-cylinder slice as visible in figure 4(d). However, in all three reconstructions the concentration changes outside the focus region are suppressed very well. This feature might prove advantageous in certain cases, for example, if the auto-fluorescence signal from other organs shall be damped.

In section 3.2 the attempt was made to link the redundancy measure to entropy and mutual information, respectively. The three quality criteria given in equations (32), (33) and (24) are very similar in their structure, though not equal. The mutual information criterion operates on the full covariance matrix while the redundancy criteria use its diagonal part only.

The MI optimization is computationally expensive due to the need of matrix decompositions, matrix inversions and the calculation of the determinant. The redundancy reduction is much more efficient as it mainly depends on the calculation of inner products of matrix rows. However, the decrease in numerical effort goes along with the neglection of the off-diagonal matrix entries.

Using the arithmetic average one could obtain even better results with respect to the condition number of the optimum sensitivity matrix. However, then also the connection to entropy minimization is lost which is why the results presented in this paper are based on the geometric averaging method.

In this paper one of the most simple stopping criteria which is the final number of sources and detectors was chosen. It is worth mentioning that the algorithm is flexible enough to include more elaborate criteria easily. For example, it could be desirable to specify the minimum required resolution or the minimum contrast-to-noise ratio. In further work also the feasibility of dynamic stopping criteria will be investigated. The principle is to calculate an initial image

quality criterion (*e.g.* the resolution). Then again optodes are iteratively dismissed from the feasible pool and as soon as the image quality decreases significantly, the algorithm stops.

6. CONCLUSION

We presented an algorithm to optimize the sensor and detector locations for fluorescence diffusion optical tomography. The possibility to bias the resultant design to be sensitive in a given area was demonstrated. In contrast to previously reported algorithms, the current formulation has a strong connection to entropy optimization.

ACKNOWLEDGMENTS

This work was supported by the project F3207-N18 granted by the Austrian Science Fund.

REFERENCES

[1] Mobley, J. and Vo-Dinh, T., "Optical properties of tissue," in [*Biomedical Photonics Handbook*], Vo-Dinh, T., ed., 2/1–2/75, CRC Press LLC, Boca Raton, USA (2003).

[2] Graves, E. E., Culver, J. P., Ripoll, J., Weissleder, R., and Ntziachristos, V., "Singular-value analysis and optimization of experimental parameters in fluorescence molecular tomography," *J. Opt. Soc. Am. A* **21**, 231–241 (2004).

[3] Lasser, T. and Ntziachristos, V., "Optimization of 360° projection fluorescence molecular tomography," *Medical Image Analysis* **11**, 389–399 (2007).

[4] Arridge, S. R., "Optical tomography in medical imaging," *Inverse Problems* **15**, R41–R93 (1999).

[5] Joshi, A., *Adaptive finite element methods for fluorescence enhanced optical tomography*, PhD thesis, Texas A&M University (2005).

[6] Schulz, R. B., Peter, J., Semmler, W., D'Andrea, C., Valentini, G., and Cubeddu, R., "Comparison of noncontact and fiber-based fluorescence-mediated tomography," *Optics Letters* **31**, 769–771 (2006).

[7] Zeidler, E., [*Applied functional analysis*], vol. 109 of *Applied Mathematical Sciences*, Springer-Verlag, New York (1995). Main principles and their applications.

[8] Gull, S. F. and Skilling, J., "Maximum entropy method in image processing," *IEE Proceedings* **131**, 646–659 (1984).

[9] Mwambela, A. J., Isaksen, O., and Johansen, G. A., "The use of entropic thresholding methods in reconstruction of capacitance tomography data," *Chemical Engineering Science* **52**, 2149–2159 (1997).

[10] Denisova, N. V., "Maximum-entropy-based tomography for gas and plasma diagnostics," *J. Phys. D: Appl. Phys.* **31**, 1888–1895 (1998).

[11] Mohammad-Djafari, A. and Demoment, G., "Maximum entropy image reconstruction in x-ray and diffraction tomography," *IEEE Transactions on Medical Imaging* **7**, 345–354 (1988).

[12] Nguyen, M. K. and Mohammad-Djafari, A., "Bayesian approach with the maximum entropy principle in image reconstruction from microwave scattered field data," *IEEE Transactions on Medical Imaging* **13**, 254–262 (1994).

[13] Ardekani, B. A., Braun, M., Hutton, B. F., Kanno, I., and Iida, H., "Minimum cross-entropy reconstruction of pet images using prior anatomical information," *Physics in Medicine and Biology* **41**, 2497–2517 (1996).

[14] Ahmed, N. A. and Gokhale, D. V., "Entropy expressions and their estimators for multivariate distributions," *IEEE Transactions on Information Theory* **35**, 688–692 (May 1989).

[15] Puntanen, S. and Styan, G. P. H., [*The Schur Complement and Its Applications (Numerical Methods and Algorithms)*], ch. Schur Complements in Statistics and Probability, 163–226, Springer (2005).

[16] Curtis, A., Michelini, A., Leslie, D., and Lomax, A., "A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys," *Geophysical Journal International* **157**, 595–606 (2004).

[17] Alexandrakis, G., Rannou, F. R., and Chatziioannou, A. F., "Tomographic bioluminescence imaging by use of a combined optical-pet (opet) system: a computer simulation feasibility study," *Physics in Medicine and Biology* **50**, 4225–4241 (2005).

[18] Keijzer, M., Star, W. M., and Storchi, P. R. M., "Optical diffusion in layered media," *Applied Optics* **27**, 1820–1824 (1988).

[19] Ntziachristos, V. and Weissleder, R., "Experimental three-dimensional fluorescence reconstruction of diffuse media by use of a normalized born approximation," *Optics Letters* **26**, 893–895 (2001).

[20] Haase, G., Kuhn, M., and Langer, U., "Parallel multigrid 3d maxwell solvers," *Parallel Computing* **27**, 761–775 (2001).

[21] Gursoy, D. and Scharfetter, H., "Optimum receiver array design for magnetic induction tomography," *IEEE Transactions on Biomedical Engineering* , In press. (2009).