IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **54** (2009) 6457–6476 doi:10.1088/0031-9155/54/21/002

A combined reconstruction–classification method for diffuse optical tomography

P Hiltunen[1], S J D Prince[2] and S Arridge[2]

[1] Department of Biomedical Engineering and Computational Science, Helsinki University of Technology, PO Box 3310, FI-02015 TKK, Finland
[2] Department of Computer Science, University College London, Gower Street London, WC1E 6B, UK

E-mail: petri.hiltunen@tkk.fi, s.prince@cs.ucl.ac.uk and s.arridge@cs.ucl.ac.uk

Received 19 March 2009, in final form 31 August 2009
Published 9 October 2009
Online at stacks.iop.org/PMB/54/6457

Abstract

We present a combined classification and reconstruction algorithm for diffuse optical tomography (DOT). DOT is a nonlinear ill-posed inverse problem. Therefore, some regularization is needed. We present a mixture of Gaussians prior, which regularizes the DOT reconstruction step. During each iteration, the parameters of a mixture model are estimated. These associate each reconstructed pixel with one of several classes based on the current estimate of the optical parameters. This classification is exploited to form a new prior distribution to regularize the reconstruction step and update the optical parameters. The algorithm can be described as an iteration between an optimization scheme with zeroth-order variable mean and variance Tikhonov regularization and an expectation-maximization scheme for estimation of the model parameters. We describe the algorithm in a general Bayesian framework. Results from simulated test cases and phantom measurements show that the algorithm enhances the contrast of the reconstructed images with good spatial accuracy. The probabilistic classifications of each image contain only a few misclassified pixels.

M This article features online multimedia enhancements

1. Introduction

Many medical imaging modalities can be described as *indirect imaging*. An image representing physical parameters is created from measurements by solving an inverse problem. Such problems may be well- or ill-posed, and linear or nonlinear. It is common for image processing techniques to be subsequently applied to these reconstructed images. A major subset of these techniques is the segmentation of structures based on the physical parameters at each pixel.

 © 2009 Institute of Physics and Engineering in Medicine

This is equivalent to *classifying* each pixel of the reconstructed image as belonging to one of a finite number of structures. For example, the image might be thresholded (Sahoo *et al* 1988, Otsu 1979), segmented using a watershed algorithm (Rettman *et al* 2002, Moga and Gabbouj 1997), clustered (Bezdek 1981, Duda and Hart 1973, Fukunaga 1972) or modelled with snakes or active shape models (Kass *et al* 1987, Staib and Duncan 1992, Cootes *et al* 1994).

These classification techniques were developed for *direct imaging* problems such as machine vision. Consequently, they are highly reliant on the quality of the reconstructed images. In this paper, we apply these ideas to indirect imaging. Specifically, we consider combining the classification and imaging steps: by considering the two steps as a single problem, we are faced with the segmentation of a small number of structures from a finite set of measurements.

Combined reconstruction–classification methods have been applied in areas such as SPECT, PET and CT (Hsiao *et al* 1998, 2002). In this paper, we apply these ideas to optical tomography, which is characterized by being both nonlinear and highly ill posed. The most relevant prior work in this modality (Guven *et al* 2005) assumed that the spatial structure of the segmentation was known (this was based on MR data), but that the physical parameters of each segment were unknown. To the best of our knowledge, no previous study has tried to simultaneously estimate both the spatial structure and the physical parameters in optical tomography.

A second way to consider our contribution is in terms of image priors. Most priors for optical tomography impose low-level information about the smoothness of the image (Paulsen and Jiang 1996, Kolehmainen *et al* 2000b, Douiri *et al* 2007, Hiltunen *et al* 2008). However, more sophisticated priors over images are commonplace in machine vision (Feng *et al* 2002, Domke *et al* 2008, Warrell and Prince 2009, Roth and Black 2005, Freeman *et al* 2000). Many of these incorporate the notion of hidden variables (Feng *et al* 2002, Warrell and Prince 2009, Freeman *et al* 2000): a complex probability distribution is induced by a generative model that depends on one or more unseen quantities.

In this paper, the reconstruction/classification model can be considered a prior with hidden variables. In common with other image priors, it imposes a probability distribution over the space of reconstructed images: it favours images in which the physical parameters are well approximated by one of a finite number of distinct values (each corresponding to one classified region). However, it is unbiased as to what these values are or how they are distributed amongst the image pixels. These quantities are held in the parameters and hidden variables of the model respectively and are determined by the data. In other words, the model favours a certain family of images, but the data themselves determine exactly which member of the family is chosen.

The structure of the paper is as follows. In section 2, we briefly introduce optical tomography. In section 3, we describe the forward model and inverse problem in optical tomography, and the reconstruction–classification algorithm. In section 4, we discuss implementation issues of the algorithm and show results from simulated and measured data.

2. Optical tomography

Diffuse optical tomography is a new medical imaging modality with potential applications in functional imaging of the brain and in breast cancer detection (Arridge 1999, Gibson *et al* 2005). This method seeks to recover optical parameters of blood and tissue from boundary measurements of light transmission in the visible and near-infrared range. The reconstructed images of the spatial distribution of tissue parameters can be related directly to physiologically important properties such as blood and tissue oxygenation state. Instrumentation for optical

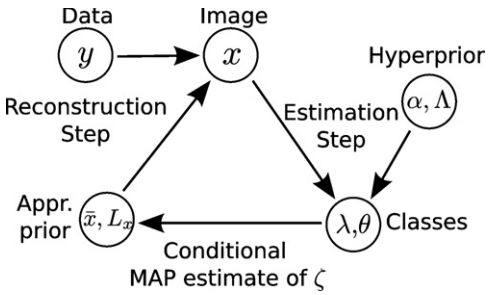

Figure 1. Overview of the approach. In the reconstruction step, we estimate the physical properties of the medium x from the data y, using a Gaussian prior with mean $\bar{x}$ based on the previous segmentation results. In the estimation step, we soft-assign pixels to clusters and estimate the properties of these clusters λ, θ. The resulting mixture model is then approximated to form the mean $\bar{x}$ and Cholesky factor of the variance L_x of the new prior. The result of the method is a reconstructed image together with class labels per pixel and statistical properties of the classes.

tomography is portable and relatively inexpensive, and can provide a viable alternative to currently available systems such as functional magnetic resonance imaging.

Data acquisition systems consist of a light source such as an infrared laser, illuminating the body surface at different source locations in succession. The light which has propagated through the tissue is then measured at multiple detector locations on the surface. Biological tissue is strongly scattering at the wavelengths used in optical tomography, which generally makes the recovery of tissue parameters from the boundary data a highly nonlinear problem.

The experimental systems in use today utilize either ultrashort input pulses (time-domain systems) or continuous intensity-modulated input (frequency domain systems). In the former case, measurements consist of the temporal dispersion of the transmitted pulse, measured at a resolution in the order of pico-seconds. In the latter case, the measurements consist of the complex intensity of the transmitted photon density wave, most commonly measured in terms of the phase shift and modulation amplitude (Nissilä *et al* 2006). The frequency domain version of the problem is considered in this paper.

3. Methods

A summary of the idea is shown in figure 1. A glossary of the notation is presented in table 2. We choose a predefined number of classes n_c. A reconstruction step consists of a few iterations of a damped Gauss–Newton optimization scheme for a regularized output least-squares problem, where the regularization scheme is first-order Tikhonov with variable mean and variance. The estimation step is an expectation-maximization (EM) method (Dempster *et al* 1977) for class labels, means and variances. The whole cycle is iterated until convergence.

Note that the EM step may be interpreted as a fuzzy K-means classification method (Mackay 2002), but we provide a more complete general hierarchical Bayesian framework that allows the imposition of priors on the mixture model parameters.

3.1. Forward model and inverse problem

The most commonly used forward model for optical diffusion tomography is the diffusion approximation (DA) to the radiative transfer equation (RTE), and is the one adopted in this paper. The implicit assumption is that the errors between this model and more accurate ones

are smaller than the measurement noise in the system, or, alternatively, can be calibrated for by analysis of the systematic errors (Arridge *et al* 2006). Under the diffusion approximation, the frequency domain version of optical tomography used in this paper is posed in terms of a complex-valued photon density $u(\mathbf{r})$ in the medium which obeys an elliptic partial differential equation:

$$-\nabla \cdot \kappa(\mathbf{r})\nabla u(\mathbf{r}) + \left(\mu_a(\mathbf{r}) + \frac{\mathrm{i}\omega}{c}\right) u(\mathbf{r}) = 0, \quad \mathbf{r} \in \Omega. \tag{1}$$

In equation (1), the optical properties are the absorption coefficient $\mu_a(\mathbf{r})$ and the reduced scattering coefficient $\mu_s'(\mathbf{r})$, governing diffuse photon transportation in the domain $\Omega \in \mathbb{R}^d$, $d = 2, 3$. The diffusion coefficient is defined as $\kappa(\mathbf{r}) = (d(\mu_a(\mathbf{r}) + \mu_s'(\mathbf{r})))^{-1}$. The speed of light in the medium is c and ω is the modulation frequency.

Equation (1) is subject to the Robin boundary condition written as

$$u(\mathbf{r}) + 2\xi\kappa(\mathbf{r})\frac{\partial u(\mathbf{r})}{\partial \mathbf{n}} = J^-(\mathbf{r}), \qquad \mathbf{r} \in \partial\Omega, \tag{2}$$

for the incoming photon current J^-, where $\mathbf{n}$ is the outward normal vector and the refractive mismatch on the boundary is taken into account via the coefficient ξ. Measurements are defined in terms of the Neumann boundary condition

$$J^+(\mathbf{r}) = -\kappa(\mathbf{r})\frac{\partial u(\mathbf{r})}{\partial \mathbf{n}}, \qquad \mathbf{r} \in \partial\Omega, \tag{3}$$

which defines the outgoing photon current J^+, and is termed *exitance* in the optical literature.

In this paper, following Schweiger *et al* (2005), the forward model is defined in terms of the logarithm of magnitude and phase of complex-valued exitance $y_{jq} = [\ln|J_q^+(\mathbf{r}_j)| \arg J_q^+(\mathbf{r}_j)]^{\mathrm{T}}$ evaluated at specific boundary locations $\mathbf{r}_j$, $j \in \{1, \ldots, n_m\}$, for each of the specified set of source locations $\mathbf{r}_q$, $q \in \{1, \ldots, n_s\}$. Similarly, the optical parameters are considered in terms of a logarithmic transformation $x_i = \ln[\mu_a(\mathbf{r}_i)\ \kappa(\mathbf{r}_i)]$, $i \in \{1, \ldots, N\}$, to ensure positivity in the image reconstruction, and are evaluated on a regular structured grid (the pixel basis) for N pixels. This leads to the model

$$y = f(x) + e\,, \qquad f : \mathbb{R}^{2N} \to \mathbb{R}^{2M} \tag{4}$$

where $M = n_m \times n_s$ is the number of measurements, and e is the assumed normally distributed additive noise

$$e = y - f(x) \sim \mathrm{N}(0, C_y).$$

where C_y is the variance of the measurement noise. This gives us a likelihood model

$$p(y|x) \propto \exp\left(-\tfrac{1}{2}(y - f(x))^{\mathrm{T}} C_y^{-1}(y - f(x))\right). \tag{5}$$

The forward model is described in more detail in Schweiger *et al* (1995), Arridge (1999), Arridge *et al* (2006).

The inverse problem in optical tomography is to solve for the optical parameters when the measured quantities and source distribution are known. In this paper, we adopt the Bayesian statistical framework, where measured data y and optical parameters x are considered as random variables. Bayes' theorem gives us the conditional posterior density of optical parameters

$$p(x|y) \propto p(y|x)p(x) \tag{6}$$

when measurement y is given (Kaipio and Somersalo 2004). In the next sections, we discuss the prior distribution for optical parameters with hyperparameters $p(x|\phi)$ and hyperpriors $p(\phi)$ and describe an iterative posterior inference algorithm for the inverse problem.

3.2. Mixture model

In most previous work, prior models (which for non-Bayesian approaches are considered as regularization penalty terms) are based, either explicitly or implicitly, on Gaussian distributions. In contrast to these models, we consider the prior model as a mixture of Gaussian distributions. To accomplish this, we introduce n_c binary variables $\zeta_{i\ell}$ at each pixel i. These indicator variables provide information about which Gaussian (class) from the mixture is present so that

$$\zeta_{i\ell} = \begin{cases} 1 & \text{pixel } i \text{ belongs to the class } \ell \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $\sum_\ell \zeta_{i\ell} = 1$ for every pixel i. We assume that indicator values at each pixel are drawn from a multinomial distribution $\zeta_i|\lambda \sim \text{Multin}(1|\lambda)$, where λ_ℓ is the overall probability of belonging to class ℓ:

$$p(\zeta_i|\lambda) = \prod_\ell \lambda_\ell^{\zeta_{i\ell}}. \tag{8}$$

The parameters λ_ℓ can be interpreted as the weights (or relative importance) of the mixture components. When the indicator variable is known, the probability distribution for the optical parameters x_i at each pixel is

$$p(x_i|\zeta_i, \theta) = \prod_\ell p(x_i|\theta_\ell)^{\zeta_{i\ell}}, \tag{9}$$

where $p(x_i|\theta_\ell)$ is a Gaussian probability density function and the parameters $\theta_\ell = \{m_\ell, C_\ell\}$ are the mean and variance of the Gaussian density.

We assume that the pixels are independently distributed so that

$$p(x, \zeta|\lambda, \theta) = p(\zeta|\lambda)p(x|\zeta, \theta) = \prod_i \prod_\ell (\lambda_\ell p(x_i|\theta_\ell))^{\zeta_{i\ell}}. \tag{10}$$

When we marginalize over the indicator variables ζ, we retrieve the *mixture of Gaussian* distribution:

$$p(x|\lambda, \theta) = \int_\zeta p(x, \zeta|\lambda, \theta)\, \mathrm{d}\zeta = \prod_i \sum_\ell \lambda_\ell p(x_i|\theta_\ell). \tag{11}$$

We also place hyperpriors over the parameters λ and θ of the mixtures of Gaussian distribution. For the mixture weights λ, the conjugate prior distribution $\lambda \sim \text{Dirichlet}(\alpha)$ is a natural choice. Parameters α_ℓ are the expected number of pixels in the class ℓ. If we set every element of α_ℓ to 1, we get a uniform distribution $p(\lambda) \propto 1$. This is a non-informative distribution indicating that we do not have any specific *a priori* information about the class probabilities. The conjugate prior distribution for the class variances is the inverse-Wishart distribution

$$p(C_\ell) \propto |C_\ell|^{-(\nu_\ell+d+1)/2} \exp\left(-\tfrac{1}{2}\text{tr}\left(\Lambda_\ell C_\ell^{-1}\right)\right), \tag{12}$$

where ν_ℓ is the number of degrees of freedom and Λ_ℓ is a scale matrix. In conjunction with the inverse-Wishart distribution, one can use a non-informative prior distribution $p(m_\ell) \propto 1$ for class means. We can also place a non-informative prior over the parameters θ of the Gaussians which takes the form of a multivariate Jeffreys density $p(\theta) \propto \prod_\ell |C_\ell|^{-(d+1)/2}$ (Gelman *et al* 2004).

3.3. Posterior inference

We use Bayes' theorem to calculate the posterior distribution of the unknown optical parameters x, mixture weights λ, Gaussian parameters θ and indicator variables ζ from

the observed measurements y:

$$p(x, \lambda, \theta, \zeta | y) \propto p(y|x)p(x|\lambda, \theta, \zeta)p(\lambda, \theta, \zeta). \tag{13}$$

Unfortunately, there is no closed form solution for the posterior distribution. Two possible approaches to this problem are to (i) calculate the maximum a posteriori (MAP) estimate of the posterior distribution or (ii) take samples from the distribution and estimate the conditional mean of the parameters. In this paper, we adopt the former approach and use a similar iterative joint MAP algorithm to Hsiao *et al* (1998). Posterior inference consists of two steps. In the *reconstruction step,* we estimate the optical properties x while fixing the parameters λ, θ, and ζ of the mixture of Gaussians prior:

$$x^{k+1} = \arg\max_x p(y|x)p(x|\zeta^k, \lambda^k, \theta^k). \tag{14}$$

In the *parameter estimation step,* we consider the optical properties x to be fixed and update the mixture model parameters λ, θ:

$$(\lambda^{k+1}, \theta^{k+1}) = \arg\max_{\lambda,\theta} p(x^{k+1}|\lambda, \theta)p(\lambda, \theta). \tag{15}$$

This step is the same as *fuzzy K-means* clustering if we use hyperprior $p(\alpha) \propto 1$ and prior $p(\theta) \propto 1$.

After parameter estimation, the MAP estimate of ζ is found from the distribution $p(\zeta|x^{k+1}, \lambda^{k+1}, \theta^{k+1})$. In the next two sections, we show how the reconstruction step transforms into a minimization problem and describe an expectation maximization (EM) approach to estimate the parameters in the estimation step.

3.3.1. Reconstruction step. In the reconstruction step, we aim to optimize the optical parameters x given the measurements y and the mixtures of Gaussians prior $p(x|\zeta, \lambda, \theta)$ using Bayes' rule as in equation (6). To solve this problem, we assume that the precision C_y^{-1} of the measurement noise can be factored as $C_y^{-1} = L_y^{\mathrm{T}} L_y$ using the Cholesky decomposition. After these changes, we can transform the maximization step (14) to a minimization problem:

$$x^{k+1} = \arg\min_x \frac{1}{2}\|L_y(y - f(x))\|^2 - \sum_i \ln \sum_\ell \lambda_\ell p(x_i|\theta_\ell). \tag{16}$$

Unfortunately, this is a non-convex optimization problem. To resolve this problem, we approximate the mixtures of Gaussian distribution at each pixel by evaluating only the cluster with maximum posterior probability. In practice, this can be done by setting the indicator variable associated with the MAP cluster to one and the others to zero so that,

$$p(x|\zeta, \lambda, \theta) = \prod_i \prod_\ell p(x_i|\theta_\ell)^{\zeta_{i\ell}} = N(\bar{x}, C_x), \tag{17}$$

where $\bar{x} \in \mathbb{R}^{2N}$ is organized in N blocks of 2 such that the ith block is equal to m_ℓ if pixel i belongs to class ℓ. Similarly $C_x \in \mathbb{R}^{2N\times 2N}$ is the block diagonal with the ith block equal to the 2×2 covariance matrix C_ℓ. The complete variance matrix is Cholesky factored as $C_x^{-1} = L_x^{\mathrm{T}} L_x$. After these operations, both the prior and likelihood terms are Gaussian and finding the MAP estimate of the optical parameters x is a convex optimization problem:

$$x^{k+1} = \arg\min_x \|L_y(y - f(x))\|^2 + \gamma\|L_x(x - \bar{x})\|^2, \tag{18}$$

where we have also included a regularization parameter γ.

3.3.2. Estimation step. In the estimation step, we estimate variables λ and θ by finding the mode of the posterior distribution $p(\lambda, \theta|x)$ where x is the current estimate from the

reconstruction step. Rather than directly maximizing the posterior as in equation (15), we maximize the log posterior so that

$$\ln p(\lambda, \theta|x) = \ln p(\zeta, \lambda, \theta|x) - \ln p(\zeta|\lambda, \theta, x). \tag{19}$$

Averaging the previous equation over ζ under the distribution $p(\zeta|\lambda^k, \theta^k, x)$, we get

$$\ln p(\lambda, \theta|x) = \mathrm{E}_k(\ln p(\zeta, \lambda, \theta|x)) - \mathrm{E}_k(\ln p(\zeta|\lambda, \theta, x)), \tag{20}$$

where the index k refers to the current iteration step. Given new estimates λ^{k+1} and θ^{k+1}, which increase $\mathrm{E}_k(\ln p(\zeta, \lambda, \theta|x))$, we also have that $\ln p(\lambda, \theta|x)$ increases because $\mathrm{E}_k(\ln p(\zeta|\lambda, \theta, x))$ reaches a maximum at the current estimates (λ^k, θ^k) (Gelman *et al* 2004).

The EM algorithm consists of two steps: the expectation- or E-step and the maximization- or M-step. In the E-step, we determine the expectation of the log posterior function $\mathrm{E}_k(\ln p(\zeta, \lambda, \theta|x))$. In the M-step, we find new estimates of λ and θ, which increase $\mathrm{E}_k(\ln p(\zeta, \lambda, \theta|x))$.

E-step: the expectation of the log posterior function simplifies to

$$\begin{aligned}\mathrm{E}_k(\ln p(\zeta, \theta, \lambda|x)) &= \int \ln p(\zeta, \theta, \lambda|x)\, p(\zeta|\theta^k, \lambda^k, x)\, \mathrm{d}\zeta \\ &= \sum_i \sum_\ell \ln(\lambda_\ell p(x_i|\theta_\ell))\mathrm{E}(\zeta_{i\ell}|\theta^k, \lambda^k, x) + \ln p(\lambda) + \ln p(\theta) - \ln p(x),\end{aligned} \tag{21}$$

where we have used the identity

$$p(\zeta, \theta, \lambda|x) = \frac{1}{p(x)} p(\lambda) p(\theta) p(x, \zeta|\lambda, \theta) \tag{22}$$

and the conditional mean of $\zeta_{i\ell}$ is

$$\begin{aligned}\mathrm{E}(\zeta_{i\ell}|\theta^k, \lambda^k, x) &= \int \zeta_{i\ell}\, p(\zeta_{i\ell}|\theta^k, \lambda^k, x)\, \mathrm{d}\zeta_{i\ell} \\ &= 0 \cdot \Pr(\zeta_{\ell i} = 0|\theta^k, \lambda^k, x) + 1 \cdot \Pr(\zeta_{i\ell} = 1|\theta^k, \lambda^k, x) \\ &= \frac{\lambda_\ell^k p\left(x_i|\theta_\ell^k\right)}{\sum_l \lambda_l^k p\left(x_i|\theta_l^k\right)}.\end{aligned} \tag{23}$$

In the following, we use the abbreviated notation $\mathrm{E}(\zeta_{i\ell}) = \mathrm{E}(\zeta_{i\ell}|\theta^k, \lambda^k, x)$. The term $\ln p(x)$ is constant during the EM algorithm and can be ignored.

M-step: we find values of the variables which maximize (21)

$$\begin{aligned}\mathrm{E}_k(\ln p(\zeta, \theta, \lambda|x)) = \sum_i \sum_\ell \mathrm{E}(\zeta_{i\ell}|\theta^k, \lambda^k, x) \left(\ln \lambda_\ell + \ln|C_\ell| - \frac{1}{2}(x_i - m_\ell)^{\mathrm{T}} C^{-1}(x_i - m_\ell)\right) \\ + \sum_\ell \left((\alpha_\ell - 1)\ln \lambda_\ell - \frac{d+1}{2}\ln|C_\ell|\right).\end{aligned} \tag{24}$$

First we maximize with respect to λ given the constraint $\sum_\ell \lambda_\ell = 1$. This can be easily done using the mode of the resulting Dirichlet distribution:

$$\lambda_\ell^{k+1} = \frac{\sum_i \mathrm{E}(\zeta_{i\ell}) + \alpha_\ell - 1}{N + \sum_\ell \alpha_\ell - n_c}, \tag{25}$$

Note that for the uninformative prior $\alpha_\ell = 1$, this reduces to

$$\lambda_\ell^{k+1} = \frac{\sum_i \mathrm{E}(\zeta_{i\ell})}{N}.$$

Secondly, the estimates of mean m_ℓ and variance C_ℓ in the case of the non-informative prior are updated:

$$m_\ell^{k+1} = \frac{\sum_i \mathrm{E}(\zeta_{i\ell}) x_i}{\sum_i \mathrm{E}(\zeta_{i\ell})}, \tag{26}$$

and

$$C_\ell^{k+1} = \frac{\sum_i \mathrm{E}(\zeta_{i\ell}) \left(x_i - m_\ell^{k+1}\right)\left(x_i - m_\ell^{k+1}\right)^{\mathrm{T}}}{\sum_i \mathrm{E}(\zeta_{i\ell}) + d + 1}. \tag{27}$$

If the inverse-Wishart distribution is used the variance update is

$$C_\ell^{k+1} = \frac{\sum_i \mathrm{E}(\zeta_{i\ell}) \left(x_i - m_\ell^{k+1}\right)\left(x_i - m_\ell^{k+1}\right)^{\mathrm{T}} + \Lambda_\ell}{\sum_i \mathrm{E}(\zeta_{i\ell}) + \nu_\ell + d + 1}. \tag{28}$$

4. Results

4.1. Implementation details

We use a finite element method for the solution of the forward problem equations (1)–(3). Details of the finite element method (FEM) model can be found in Arridge *et al* (1993), Schweiger *et al* (1995), Schweiger and Arridge (1997). The implementation uses a library of C++ classes together with a Matlab interface (Schweiger and Arridge 2008). The EM algorithm is written in Matlab.

An outline of the algorithm is listed in algorithm 1 and contains nested iterations. The outer iteration consists of alternating reconstruction and estimation steps. There are two inner iterations: (i) a damped Gauss–Newton minimization algorithm with inexact line search (Schweiger *et al* 2005) and (ii) an EM algorithm.

To ensure convergence in the reconstruction step, data scaling in the likelihood term was used. The logarithm of the amplitude and phase were scaled by the norms of their initial errors respectively (Schweiger *et al* 2005). This is a standard technique in optimization problems where the data have elements with different scales.

In practice, this is done by rescaling the Cholesky decomposition of the noise measurement variance C_y,

$$\tilde{L}_y = \begin{bmatrix} \frac{L_{\ln A}}{s_c^{\ln A}} & 0 \\ 0 & \frac{L_\phi}{s_c^\phi} \end{bmatrix}, \tag{29}$$

where the scaling coefficients $s_c^m = \|L_m(y_m - f_m(x^0))\|^2$ and x^0 are the initial estimates of the optical parameters.

In the literature related to Tikhonov regularization, there are several suggested methods to find reasonable values of the coefficient γ (Hansen and O'Leary 1993, Vogel 2002). We have selected γ which gave reasonable results by visual inspection.

Algorithm 1 Overview of the reconstruction–segmentation algorithm. Note that the outer iteration is continued until the reconstruction and estimation steps do not produce further changes.

Initialize parameters
while not converged **do**
  $k = 0$
  Reconstruction step, equation (14)
    while $k <$ max. iterations or change in x is small
      Calculate Jacobian matrix J of $f(x^k)$
      Solve $\left(J^{\mathrm{T}} L_y^{\mathrm{T}} L_y J + \gamma L_x^{\mathrm{T}} L_x\right)\delta x^k = J^{\mathrm{T}} L_y^{\mathrm{T}}(y - f(x^k)) - \gamma L_x(x^k - \overline{x})$
      Do inexact line search in direction δx^k
      Update x^{k+1}
    end while
  Set $x = x^k$
  $k = 0$
  Estimation step, equation (15)
    while $k <$ max. iterations or change in $\mathrm{E}_k(\ln p(\zeta, \theta, \lambda|x))$ is small
      E-Step
        Determine equation (21)
      M-Step
        Solve λ_ℓ^{k+1} from equation (25)
        Solve m_ℓ^{k+1} from equation (26)
        Solve C_ℓ^{k+1} from equation (27) or (28)
      $k = k + 1$
    end while
  Set $\lambda_\ell = \lambda_\ell^k$, $C_\ell = C_\ell^k$, and $m_\ell = m_\ell^k$
  Recalculate $\bar{x}$ and L_x
end while

4.2. Simulated data

We tested our algorithm with simulated data from a two-dimensional object. The object is a circle of diameter 5 cm. There are three inclusions of diameter 1 cm each depicted as dashed lines in figures 2 and 3. Two of the inclusions had the same absorption coefficient and two had the same diffusion coefficient. The optical values and class labels are listed in table 1. The refractive index was 1.4 for both background and inclusions.

There are 32 sources and 32 detectors placed at equal distances around the object. Simulated data were generated from a FEM approximation with an unstructured mesh of 13 051 nodes and 25 752 triangular elements. 1% multiplicative white noise was added to the data. The FEM approximation in the forward problem used a mesh with 3511 nodes and

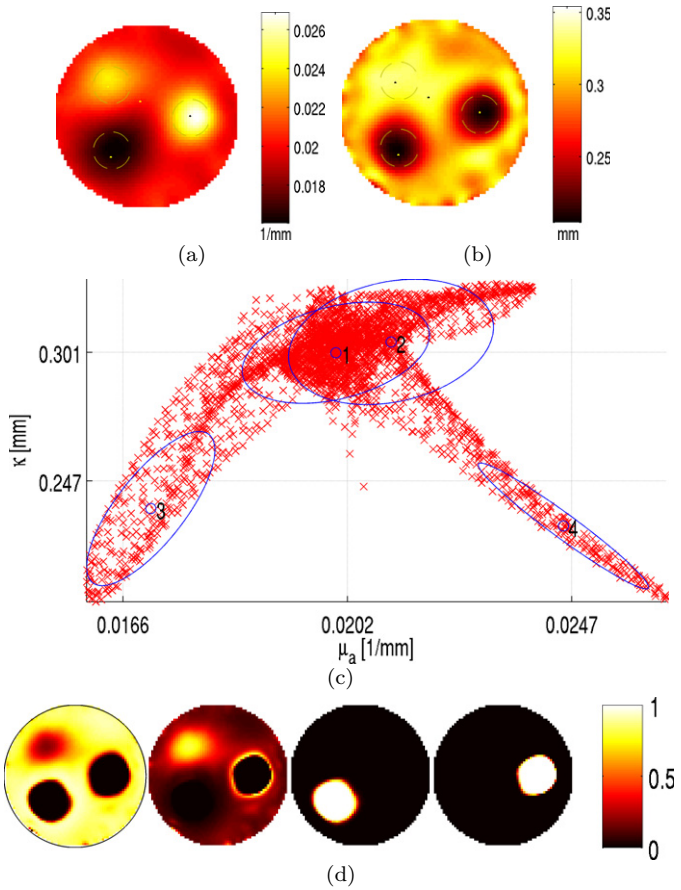

Figure 2. Results after the first reconstruction and estimation step. (a) Absorption coefficient. (b) Diffusion coefficient. (c) Scatter plot of optical parameters in $(\ln \mu_a, \ln \kappa)$-space. (d) Probabilistic classification of every pixel (classes 1–4 from left to right). Dots in (a) and (b) present selected initial values of the class means in the estimation step.

Table 1. Optical properties used in the simulated object.

	μ_a (mm^{-1})	κ (mm)	Class label
Background	0.02	0.3	1
Inclusion 1	0.03	0.4	2
Inclusion 2	0.01	0.15	3
Inclusion 3	0.03	0.15	4

6840 triangular elements in an unstructured grid and optical parameters were discretized in a structured grid which contained 3096 pixels.

We set starting values of optical parameters $x_0 = \ln[0.02\ 0.3]$. In first estimation step, variances were initialized to $C_\ell = 10^{-2}I$ and values of the inv-Wishart parameters were $\nu_\ell = 1$ and $\Lambda_\ell = 10^{-3}I$. Initial means in the first estimation step were selected from the first reconstruction result. Selected means are presented as dots in figures 2(a) and (b). The value of the variable γ was 10^{-4}. The maximum number of reconstruction steps was limited to five steps and estimation steps to one step within each outer iteration.

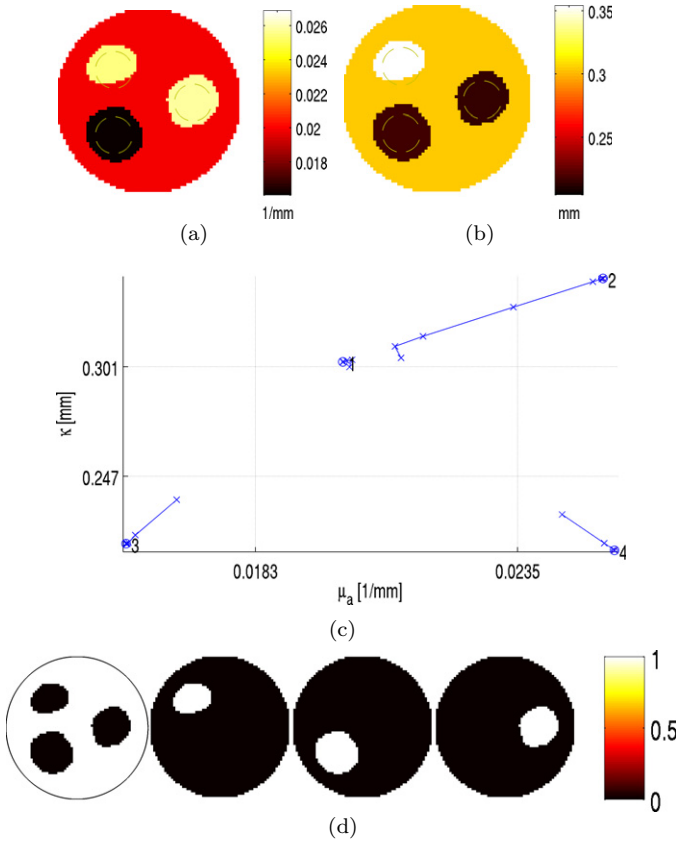

Figure 3. Results after ten reconstruction and estimation steps when the algorithm has reached the maximum number of steps; (a)–(d) are the same as in figure 2. Tails with × in (c) represent the convergence of the class means.

The result after the first reconstruction and estimation step is presented in figure 2. We have used a non-informative hyperprior ($\alpha = 1$) for λ. Reconstructed absorption and diffusion coefficients are presented in figures 2(a) and (b), respectively. The algorithm was initialized with all the class labels set to the background class. Hence, the first step of reconstruction uses a unimodal Gaussian prior with initial optical parameters as mean and results in blurred images. This can be seen in figure 2(c) where the optical parameters of every pixel are plotted in $(\ln \mu_a, \ln \kappa)$-space. The small circles and ellipses represent the estimated mean and variance of classes respectively. In figure 2(d), we depict the conditional probability of belonging to class ℓ $\Pr(\zeta_{i\ell} = 1|\theta^k, \lambda^k, x^k)$ for every pixel i. This is sometimes referred to as 'fuzzy labelling'.

Figure 3 shows the same information as figure 2 after ten reconstruction and estimation steps when the algorithm has reached the maximum number of iterations. Now absorption (figure 3(a)) and diffusion (figure 3(b)) coefficients have sharp edges near the class boundaries and the scatter plot (figure 3(c)) is clearly separated. In the scatter plot, the lines marked with crosses and ending in ellipses represent the trajectories of the label means. The complete iterative steps are shown in the animations associated with this paper (movie 1 available at stacks.iop.org/PMB/54/6457).

Table 2. Definition of variables.

Quantity	Meaning	Type
$\zeta_{i\ell}$	Indicator of class membership per pixel	Binary
n_c	Number of classes	Integer
$N = \sum_{i\ell} \zeta_{i\ell}$	Number of pixels in the image	Integer
n_s	Number of sources	Integer
n_m	Number of detectors	Integer
M	Number of complex measurements	Integer
λ_ℓ	Fraction of pixels in class ℓ	Real $\in [0, 1]$
m_ℓ	Mean of class ℓ	Real 2-vector
C_ℓ	Variance of class ℓ	Real 2×2 matrix
θ_ℓ	Parameters for pdf for class ℓ	$\{m_\ell, C_\ell\}$
α_ℓ	Parameter of Dirichlet distribution	Positive real
ν_ℓ	Parameter of inv-Wishart distribution	Positive real
Λ_ℓ	Parameter of inv-Wishart distribution	Real 2×2 matrix
γ	Regularization scaling parameter	Positive real
J^-	Source current	Complex function
J^+	Exitance	Complex function
y	Combined amplitude and phase data	Real $2M$-vector
C_y	Variance of measurement noise	Real $2N \times 2N$ matrix
x	Combined image of μ_a and κ	Real $2N$-vector
$\overline{x}$	Mean of image prior	Real $2N$-vector

We have estimated the probabilistic classification error using

$$\mathcal{E}_{\text{classification}} = \frac{\sum_i p(\text{pixel } i \text{ misclassified})}{N}. \tag{30}$$

After the first step, the classification error was 26.5% and it converged to 11.7%.

4.3. Noise analysis

In this section, we compare the reconstruction–classification algorithm of section 3 to a 'conventional approach' in which the regularized Guass–Newton algorithm is run to convergence with a fixed zero-order Tikhonov regularization term (i.e. equation (18), where $\bar{x} = x^0$ and $L_x = I$), followed by running the estimation step to convergence (maximum number of iterations was 20). In each case, the algorithms were run with the same noise instantiations, initial conditions and initial seed points for multiple trials, and the statistics of the reconstructed images and classifications are compared. The number of trials used was 25. We used two different methods to select the regularization level for the conventional algorithm: (i) the choice $\gamma = 0.0056$ gave optimal results in the sense of minimizing the reconstruction error for a known ground truth; i.e. $\|x - x_{\text{true}}\|$, and (ii) the choice $\gamma = 5.6 \times 10^{-4}$ gave optimal results in the sense of the classification error (30).

From the set of 25 trials, we determined the mean and standard deviation of the classification images and the mean and standard deviation of the bias ($\sqrt{E[(x - x_{\text{true}})^2]}$). The mean of the reconstructions and standard deviation of the bias are shown in figure 4 and the mean and the standard deviation of the classifications are shown in figure 5.

These results shows that our reconstruction–classification method and the conventional method with regularization level (i) have similar total variance but the reconstruction–classification method generates sharper images and lower classification error. Using

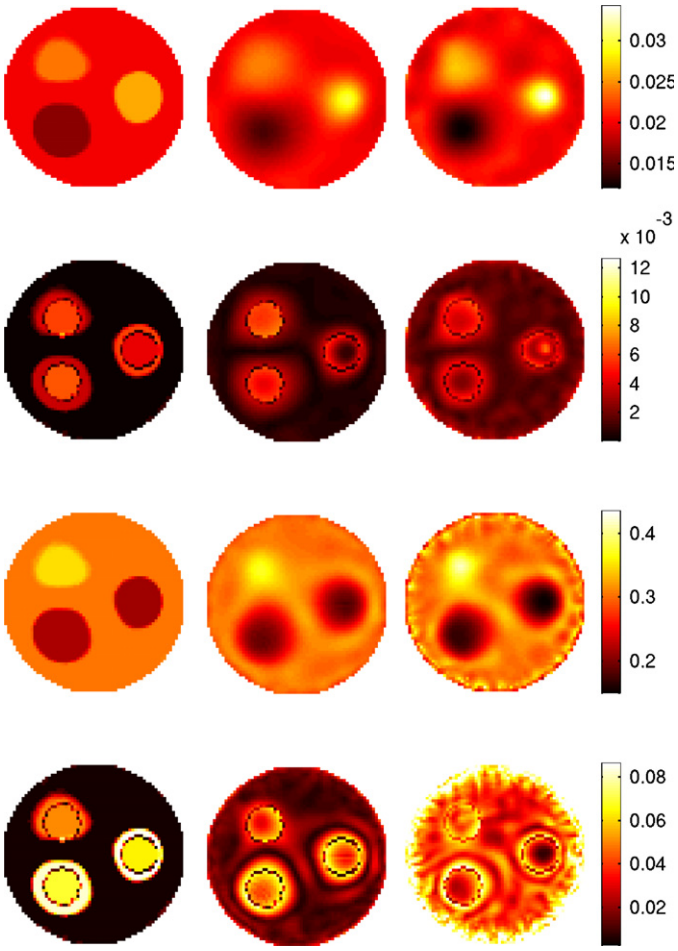

Figure 4. Mean of the absorption (first row) and the diffusion (third row) parameters from 25 reconstructions with different algorithms: presented algorithm (left column), Tikhonov regularization $\gamma = 0.0056$ (middle column) and Tikhonov regularization $\gamma = 5.6 \times 10^{-4}$ (right column). Standard deviation of the bias ($\sqrt{E[(x - x_{\text{true}})^2]}$) for the absorption parameter (second row) and the diffusion parameter (fourth row).

regularization level (ii) leads to greater error in the reconstructions. Total variance of the bias for the different methods was: absorption 0.017 and diffusion 3.25 in the reconstruction–classification algorithm, absorption 0.014 and diffusion 2.48 in case (i), and absorption 0.020 and diffusion 12.0 in case (ii). For the reconstruction–classification algorithm, the variance of the classification is almost all contributed by small errors in the localization of the edges of the objects, whereas for the conventional algorithm the variance is distributed throughout the images. The mean classification error for the reconstruction–classification algorithm was 12.8%, whereas for the conventional algorithm, it was 17.6% for regularization level (i) and 15.1% for regularization level (ii).

4.4. Phantom measurement

For experimental data, we used the phantom measurements described in Nissilä *et al* (2006). The phantom had a height of 110 mm and a diameter of 69.25 mm (see figure 6). Data y_{hmg}

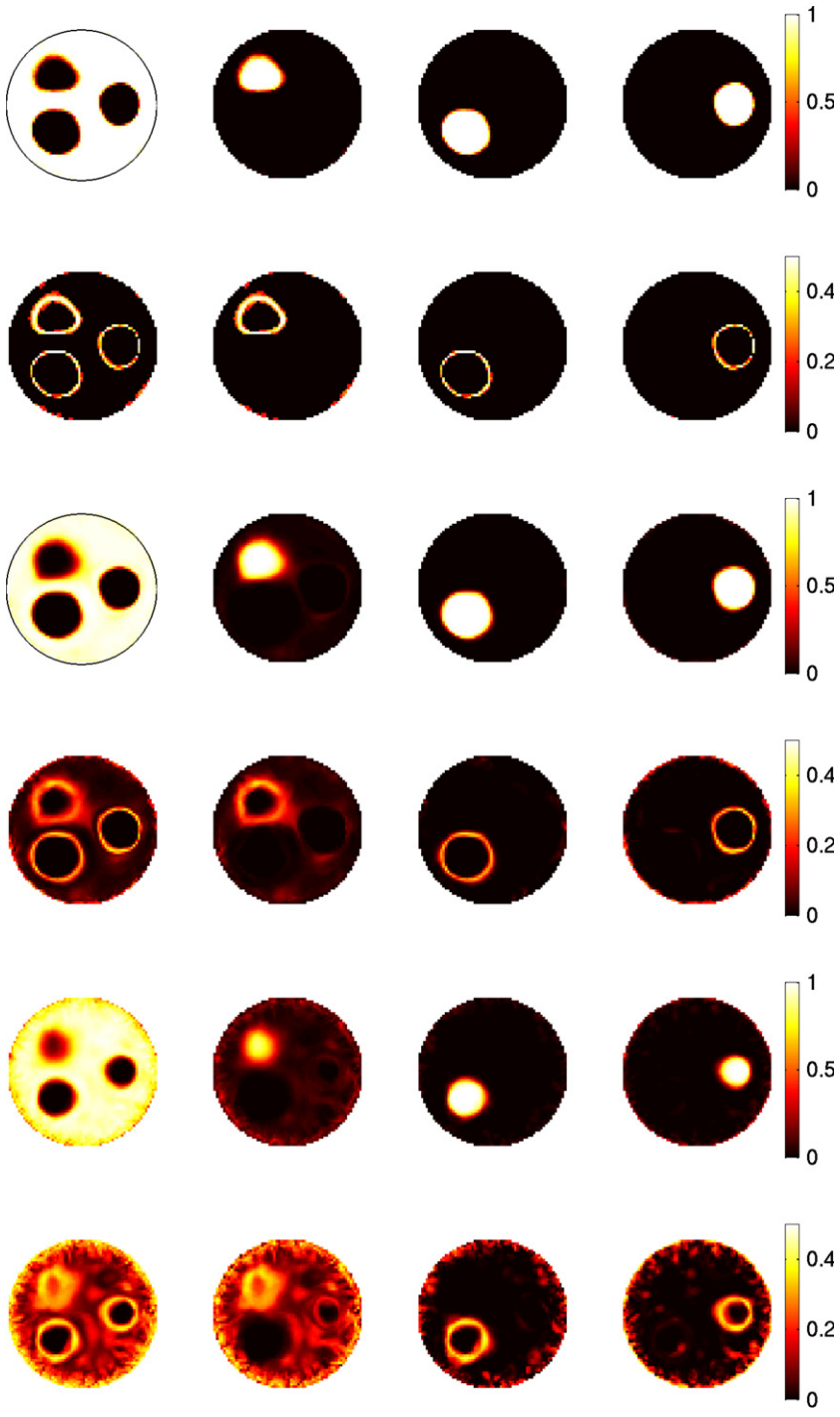

Figure 5. Mean of the probabilistic classification from 25 reconstructions with different algorithms: resented algorithm (first row), Tikhonov regularization $\gamma = 0.0056$ (second row) and Tikhonov regularization $\gamma = 5.6 \times 10^{-4}$ (third row). Standard deviation of the probabilistic classification is in the second, fourth and sixth rows, respectively. Classes from 1 to 4 are in columns from 1 to 4.

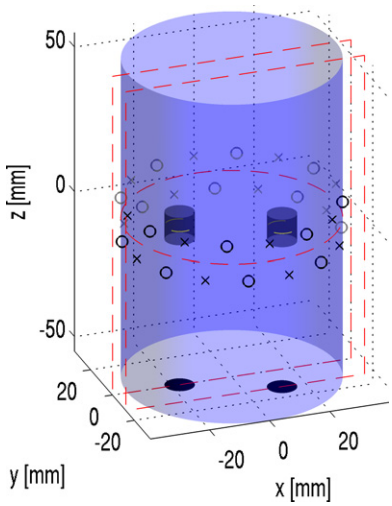

Figure 6. Geometry of the non-homogeneous phantom.

from a homogeneous phantom and data y_{inc} from a phantom with inclusions were measured and used to generate difference data. The inclusions both had cylindrical shape, diameters of 9.5 mm and heights of 9.5 mm. The data used for reconstruction were based on the difference of measurements with the homogeneous forward model added $y \equiv f(x_{\text{hmg}}) + y_{\text{inc}} - y_{\text{hmg}}$. The background optical properties were $\mu_a \approx 0.0097$ mm^{-1} and $\mu_s' \approx 1.04$ mm^{-1} and the refractive index was 1.56. One inclusion had approximately twice the background absorption coefficient and the other had approximately twice the background reduced scattering coefficient. More details of the measurement and phantoms can be found in Nissilä *et al* (2006).

The instrumentation had 16 sources and 16 detectors. Sources and detectors were placed in two rings around the phantom such that half of the sources and half of the detectors were in the upper ring and the other halves in the lower ring (see figure 6). Source detector separations less than 30 mm were not used. Therefore, 192 measurements of amplitude and phase were acquired. The FEM approximation of the forward problem was done in an unstructured grid which contained 31 171 nodes and 21 207 tetrahedral elements and optical parameters were discretized in a structured grid which contained 23 541 voxels. The reconstruction and estimation were done in $(\ln \mu_a, \ln \kappa)$-space and the reduced scattering coefficient was solved afterwards.

In the phantom measurement algorithm was initialized as in the simulated case except $x_0 = \ln[0.0097\ 0.32]$. Selected means are presented as dots in figures 7(a) and (b). In estimation step, we used only voxels approximately between $z = [48\ 62]$ (inclusion are between $z = [50.2\ 59.8]$). The value of the variable γ was 10^{-3}. In first estimation step, variances were initialized to $C_\ell = 10^{-3}I$. The values of the inv-Wishart parameters were $\nu_\ell = 2000$ and $\Lambda_\ell = 0.5I$ for background class. For other two classes, parameters were $\nu_\ell = 200$ and $\Lambda_\ell = 0.05I$. Initial means in the first estimation step were selected from the first reconstruction result. Selected means are presented as dots in figures 2(a) and (b). The value of the variable γ was 10^{-6}. The maximum number of reconstruction steps was limited to ten steps and estimation steps to one step.

Figure 7 shows the reconstructed optical parameter from two different planes when the algorithm has reached maximum ten iterations. The reconstructed images show some crosstalk

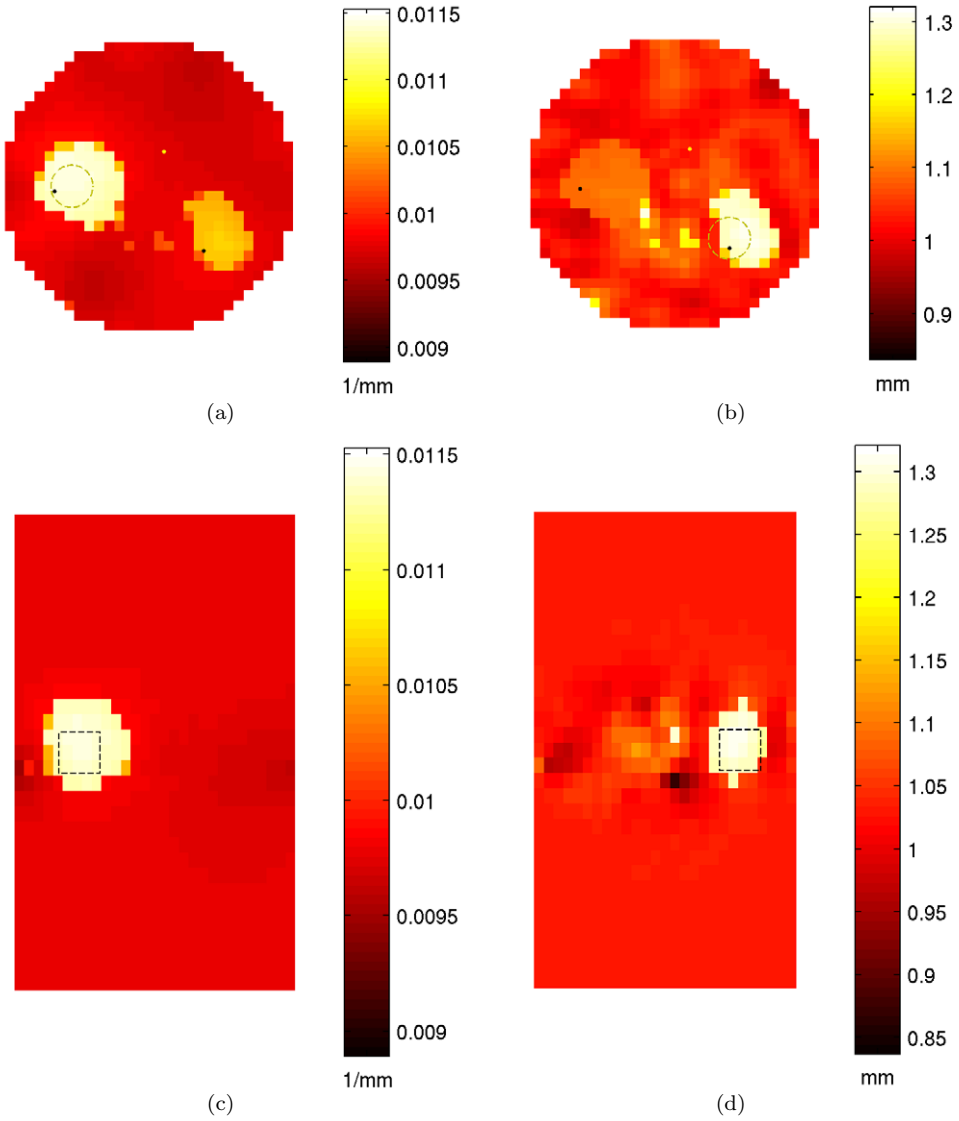

Figure 7. Reconstructed (a) absorption coefficient from plane $z = 0$, (b) reduced scattering coefficient from plane $z = 0$, (c) absorption coefficient from plane $y = 0$, and (d) reduced scattering coefficient from plane $y = -12.24$ after ten reconstruction and estimation steps when the algorithm has reached the maximum value of the iterations. Dashed lines present true locations of the inclusions. Dots in (a) and (b) present selected initial values of the class means in the estimation step.

between the estimates of absorption and scattering. From figure 8, we can see that the contrast has substantially improved during iteration, but crosstalk is still present. There are also a few misclassified voxels which are not near the inclusions or the boundary (see figure 8(b)). The complete iterative steps are shown in the animations associated with this paper (movie 2 available at stacks.iop.org/PMB/54/6457).

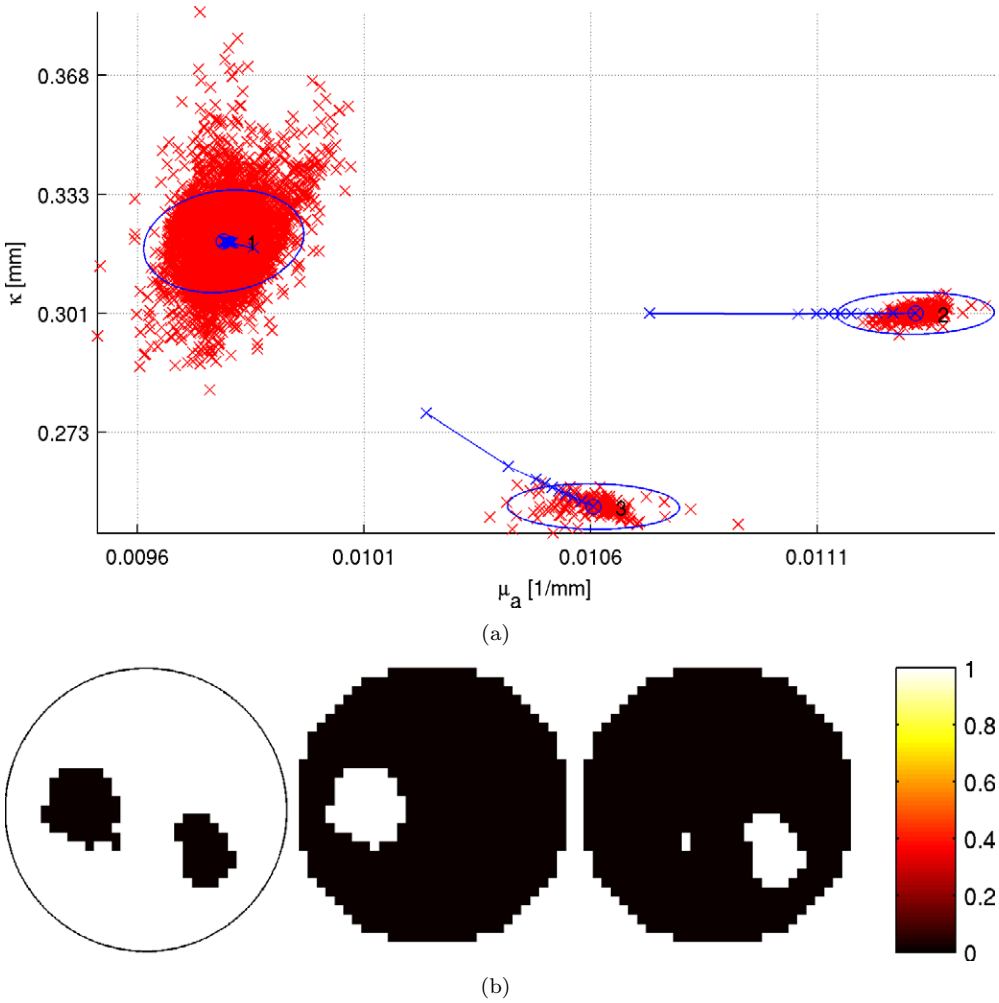

Figure 8. (a) Scatter plot of the absorption and diffusion coefficient when the algorithm has converged. Tails with × represent the convergence of the class means. (b) Probabilistic classification of the classes

5. Discussion

In this paper we have presented a simultaneous reconstruction and classification algorithm for optical tomography. We have adopted the Bayesian framework which described a statistical model of the optical parameters. In our model, optical parameters are assumed to follow the Gaussian mixture model with unknown coefficients. We have presented an algorithm which successively estimates the optical parameters and the mixture model parameters. Estimating the optical parameters simplifies to Tikhonov regularized reconstruction with variable mean and variance, and the estimates of the mixture model are found using the EM algorithm. Results from simulated data and phantom measurements showed that our algorithm can accurately classify inclusions and background and also enhance the contrast of the reconstructed optical parameters.

Our alternating reconstruction and segmentation algorithm tries to separate optical parameters into nearly homogeneous regions. The method uses purely statistical information and without any knowledge of the spatial structure. Prior knowledge of the expected proportion of the pixels in each class could have been incorporated via the hyperparameter α but was not used in the examples presented here.

The reconstruction–classification method we have presented may be compared with shape-based methods which assume that the reconstructed images consist of well-defined geometric regions and the boundaries between them (Kolehmainen *et al* 1999, Arridge *et al* 2006, 2008, Zacharopoulos *et al* 2006). It may also be compared to level set methods which allow for topological changes of the structures and regularization based on shape metrics (Schweiger *et al* 2006, 2008). However, both of these methods assume uniform contrast levels within each region that are usually assumed to be known *a priori*. In cases where these are treated as unknowns they typically exhibit a degree of cross-talk between object volume and contrast (Kolehmainen *et al* 2000a).

In comparison to the algorithm described in Guven *et al* (2005) our method produces classification at each pixel rather than assuming this based on auxiliary information. A further difference is that our implementation does not update the regularization parameter γ during iteration. We selected γ by visual inspection of the reconstructed optical parameters. An alternative approach would have been to treat γ as an unknown variable within the Bayesian framework.

We have limited our examples to the case where the number of the Gaussian distributions in the mixture model is known. We could consider two different approaches to relax this limitation. First, we can include one class with large variance and keep that fixed in the estimation step. This class can be interpreted as an outlier class which collects pixels which do not fit any other class. Alternatively, we can apply methods for automatically estimating the number of components in a mixture model. The most desirable approach would be to fit models with different numbers of components Bayesian Model Comparison (Mackay 2002) to find which is most probable. Unfortunately, this requires estimating a joint posterior distribution over both the hidden variables and the parameters of the mixture model and this is intractable. One approach is to use a variational approach to approximate this posterior (Bishop 2006). A second method would be to use a Monte Carlo method to sample from a Dirichlet Process mixture model and to choose the modal number of classes (Neal 2000). While it would be possible to apply these methods here, they are quite complex and beyond the scope of this article.

Our current algorithm does not assume any spatial correlation between neighbouring pixels. This can lead to small groups of isolated pixels appearing spatially distant from the main structure constituting a particular class, as seen for example in figure 8(b) where one of the inclusions shows some pixels included which belong in the background class. In future work, we will focus on investigating models which include neighbouring information and on building models where prior knowledge of class probabilities is available, for example using multimodality images.

Acknowledgments

The authors would like to thank D.Sc Ilkka Nissilä for providing data, and the CSC-IT Center for Science Ltd for computer resources. The research of Petri Hiltunen was supported by the Emil Aaltonen Foundation. Professor Arridge and Dr Prince acknowledge support from EPSRC grant EP/E034950/1.

References

Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** R41–93

Arridge S R, Dorn O, Kaipio J P, Kolehmainen V, Schweiger M, Tarvainen T, Vauhkonen M and Zacharopoulos A 2006 Reconstruction of subdomain boundaries of piecewise constant coefficients of the radiative transport equation from optical tomography data *Inverse Problems* **22** 2175–98

Arridge S R, Dorn O, Kolehmainen V, Schweiger M and Zacharopoulos A 2008 Parameter and structure reconstruction in optical tomography *Proc. of 6th Int. Conf. on Inverse Problems in Engineering: Theory and Practice, J Phys.: Conf. Ser.* **125** 012001

Arridge S R, Kaipio J, Kolehmainen , V Schweiger M, Somersalo E, Tarvainen T and Vauhkonen M 2006 Approximation errors and model reduction with an application in optical diffusion tomography *Inverse Problems* **22** 175–96

Arridge S R, Schweiger M, Hiraoka M and Delpy D T 1993 A finite element approach for modeling photon transport in tissue *Med. Phys.* **20** 299–309

Bezdek L 1981 *Pattern Recognition with Fuzzy Objective Function Algorithm* (New York: Plenum)

Bishop C M 2006 *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Berlin: Springer)

Cootes T F, Hill A, Taylor C J and Haslam J 1994 Use of active shape models for locating structures in medical images *Image Vis. Comput.* **25** 355–66

Dempster A, Laird N and Rubin D 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc.* B **39** 1–38

Domke J, Karapurkar A and Aloimonos Y 2008 Who killed the directed model? *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2008) (24–26 June 2008, Anchorage, AL, USA)*

Douiri A, Schweiger M, Riley J and Arridge S R 2007 Anisotropic diffusion regularisation methods for diffuse optical tomography using edge prior information *Meas. Sci. Technol.* **18** 87–95

Duda R and Hart P 1973 *Pattern Classification and Scene Analysis* (New York: Wiley)

Feng X, Williams C and Felderhof S 2002 Combining belief networks and neural networks for scene segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 467–83

Freeman W T, Pasztor E C and Carmichael O T 2000 Learning low-level vision *Int. J. Comput. Vision* **40** 25–47

Fukunaga K 1972 *Introduction to Statistical Pattern Recognition* (New York: Academic)

Gelman A, Carlin J B, Stern H S and Rubin D B 2004 *Bayesian Data Analysis* (Boca Raton, FL: Chapman and Hall/CRC)

Gibson A, Hebden J and Arridge S R 2005 Recent advances in diffuse optical tomography *Phys. Med. Biol.* **50** R1–43

Guven M, Yazici B, Intes X and Chance B 2005 Diffuse optical tomography with *a priori* anatomical information *Phys. Med. Biol.* **50** 2837–58

Hansen P C and O'Leary D P 1993 The use of the L-curve in the regularization of discrete ill-posed problems *SIAM J. Sci. Comput.* **14** 1487–503

Hiltunen P, Calvetti D and Somersalo E 2008 An adaptive smoothness regularization algorithm for optical tomography *Opt. Express* **16** 19957–77

Hsiao I-T, Rangarajan A and Gindi G 1998 Joint-MAP reconstruction/segmentation for transmission tomography using mixture-models as priors *Proc. IEEE Nuclear Science Symp. and Medical Imaging Conf.* vol 3 pp 1689–93

Hsiao I-T, Rangarajan A and Gindi G 2002 Joint-MAP Bayesian tomographic reconstruction with a gamma-mixture prior *IEEE Trans. Image Process.* A **11** 1466–77

Kaipio J and Somersalo E 2004 *Statistical and Computational Inverse Problems* (Berlin: Springer)

Kass M, Witkin A and Terzopoulos D 1987 Snakes: active contour models *Int. J. Comput. Vis.* **1** 321–31

Kolehmainen V, Arridge S R, Lionheart W R B, Vauhkonen M and Kaipio J P 1999 Recovery of region boundaries of piecewise constant coefficients of an elliptic PDE from boundary data *Inverse Problems* **15** 1375–91

Kolehmainen V, Arridge S R, Vauhkonen M and Kaipio J P 2000a Simultaneous reconstruction of internal tissue region boundaries and coefficients in optical diffusion tomography *Phys. Med. Biol.* **45** 3267–83

Kolehmainen V, Vauhkonen M, Kaipio J P and Arridge S R 2000b Recovery of piecewise constant coefficients in optical diffusion tomography *Opt. Express* **7** 468–80

Mackay D J C 2002 *Information Theory, Inference & Learning Algorithms* (Cambridge: Cambridge University Press)

Moga A and Gabbouj M 1997 Parallel image component labeling with watershed transformation *IEEE Trans. Pattern. Anal. Mach. Intell.* **19** 441–50

Neal R M 2000 Markov chain sampling methods for Dirichlet process mixture models *J. Comput. Graph. Stat.* **9** 249–65

Nissilä I *et al* 2006 Comparison between a time-domain and a frequency-domain system for optical tomography *J. Biomed. Opt.* **11** 064015

Otsu N 1979 A threshold selection method from gray-level histograms *IEEE Trans. Syst. Man Cybern.* **9** 62–6

Paulsen K D and Jiang H 1996 Enhanced frequency-domain optical image reconstruction in tissues through total variation minimization *Appl. Opt.* **35** 3447–58

Rettman M, Han X and Prince J L 2002 Automated sulcal segmentation using watersheds on the cortical surface *NeuroImage* **15** 329–44

Roth S and Black M J 2005 Fields of experts: a framework for learning image priors *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR '05) (20–26 June 2005)* vol 2 pp 860–7

Sahoo P, Soltani S, Wong A K C and Chen Y C 1988 Survey of thresholding techniques *Comput. Vis. Graph. Image Process.* **41** 233–60

Schweiger M and Arridge S R 1997 The finite element model for the propagation of light in scattering media: frequency domain case *Med. Phys.* **24** 895–902

Schweiger M and Arridge S R 2008 Toast reconstruction package http://web4.cs.ucl.ac.uk/research/vis/toast/

Schweiger M, Arridge S R, Dorn O, Zacharopoulos A and Kolehmainen V 2006 Reconstructing absorption and diffusion shape profiles in optical tomography using a level set technique *Opt. Lett.* **31** 471–3

Schweiger M, Arridge S R, Hiraoka M and Delpy D T 1995 The finite element model for the propagation of light in scattering media: boundary and source conditions *Med. Phys.* **22** 1779–92

Schweiger M, Arridge S R and Nissilä I 2005 Gauss–Newton method for image reconstruction in diffuse optical tomography *Phys. Med. Biol.* **50** 2365–86

Schweiger M, Dorn O and Arridge S R 2008 3-d shape and contrast reconstruction in optical tomography with level sets *Proc. 4th AIP Int. Conf. and the 1st Congress of the IPIA, J. Phys.: Conf. Ser.* **124** 012043

Staib L and Duncan J 1992 Boundary finding with parametrically deformable models *IEEE Trans. Pattern Anal. Mach. Intell.* **14** 1061–75

Vogel C R 2002 *Computational Methods for Inverse Problems* (Philadelphia, PA: SIAM)

Warrell J, Prince S J D and Moore A 2009 Epitomized priors for multi-labeling problems *CVPR 2009*

Zacharopoulos A, Arridge S R, Dorn O, Kolehmainen V and Sikora J 2006 Three dimensional reconstruction of shape and piecewise constant region values for optical tomography using spherical harmonic parameterisation and a boundary element method *Inverse Problems* **22** 1509–32