


An Automated Preprocessing Method for Diffuse Optical Tomography to Improve Breast Cancer Diagnosis

Technology in Cancer Research & Treatment
 Volume 17: 1-12
 © The Author(s) 2018
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/1533033818802791
journals.sagepub.com/home/tct


Murad Althobaiti, PhD¹, Hamed Vavadi, PhD², and Quing Zhu, PhD³

Abstract

The ultrasound-guided diffuse optical tomography is a noninvasive imaging technique for breast cancer diagnosis and treatment monitoring. The technique uses a handheld probe capable of providing measurements of multiple wavelengths in a few seconds. These measurements are used to estimate optical absorptions of lesions and calculate the total hemoglobin concentration. Any measurement errors caused by low signal to noise ratio data and/or movements during data acquisition would reduce the accuracy of reconstructed total hemoglobin concentration. In this article, we introduce an automated preprocessing method that combines data collected from multiple sets of lesion measurements of 4 optical wavelengths to detect and correct outliers in the perturbation. Two new measures of correlation between each pair of wavelength measurements and a wavelength consistency index of all reconstructed absorption maps are introduced. For phantom and patients' data without evidence of measurement errors, the correlation coefficient between each pair of wavelength measurements was above 0.6. However, for patients with measurement errors, the correlation coefficient was much lower. After applying the correction method to 18 patients' data with measurement errors, the correlation has improved and the wavelength consistency index is in the same range as the cases without wavelength-dependent measurement errors. The results show an improvement in classification of malignant and benign lesions.

Keywords

diffuse optical tomography, ultrasound, optical imaging reconstruction, breast cancer

Abbreviations

DOT, diffuse optical tomography; MNR, maximum normed residual; NIR, near-infrared; PMT, photomultiplier; SNR, signal to noise; tHb, total hemoglobin concentration; US, ultrasound

Received: March 11, 2018; Revised: June 17, 2018; Accepted: August 24, 2018.

Introduction

Diffuse optical tomography (DOT) is a noninvasive medical imaging technique capable of providing functional information of the tissue. This technology utilizes light in the near-infrared (NIR) window to assess the interior optical parameters of the tissue, which are closely related to the oxygenated and deoxygenated hemoglobin content within the tissue.¹⁻⁵ Because of its high sensitivity and low cost, DOT applications are growing in many areas, such as functional brain imaging, breast cancer detection and treatment monitoring, and many others.⁶⁻¹¹

The intense NIR light scattering in tissue imposes a significant challenge to DOT on its spatial resolution and localization accuracy. Coregistration approaches of DOT with high-resolution imaging modalities, such as X-ray, magnetic resonance imaging, or ultrasound (US), have been investigated.^{9,12-15} In

these approaches, a lesion is localized and the information of lesion size, shape, and depth obtained from a high-resolution imaging modality can be used to guide the DOT image reconstruction.¹⁶⁻²⁰ The US-guided DOT approach developed by our

¹ Biomedical Engineering Department, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

² Biomedical Engineering Department, University of Connecticut, Mansfield, CT, USA

³ Biomedical Engineering Department, Washington University in St Louis, St Louis, MO, USA

Corresponding Author:

Quing Zhu, PhD, Biomedical Engineering Department, Washington University in St Louis, St Louis, MO 63130, USA.
 Email: zhu.q@wustl.edu



group has improved lesion location uncertainty and reconstruction accuracy of optical parameters.²¹⁻²³ In recent years, a field of shape-based reconstruction techniques has emerged where a priori information about the topology, the approximate location, and shapes of the unknown subdomains and their optical properties can be incorporated in the inversion directly.²⁴ In the US-guided DOT approach, a semiautomated level set method that extracts tumor information from US images for DOT image reconstruction was introduced.²⁵

Previous studies have shown the capacity of US-guided DOT in classifying malignant lesions from benign masses.^{26,27} However, there are number of challenges for the wide clinical acceptance of this technology. First, the contralateral reference breast measurements used to compute weight matrix and perturbation, which is the normalized difference between lesion and reference measurements, may not be homogeneous. This can reduce the accuracy of fitted background optical properties and then computed weight matrix, which in turn causes inaccuracy in reconstructed lesion optical properties. Second, measurement errors due to low signal to noise ratio (SNR) data at longer source and detector distances and bad coupling between skin and breast tissue are wavelength dependent. For example, SNR of wavelength 740 nm at longer source and detector distances is much lower than longer wavelengths due to higher light absorption of darker skin and skin pigment.²⁸⁻³⁰ Additionally, our DOT system uses a handheld probe which is placed on top of patient's breast while the patient is in a supine position. Movements of operator's hand or patient may lead to a bad coupling between the light guides and the breast, which may result in outliers in measurements and then in the calculated perturbations. These outlier measurements are random and significantly affect the accuracy of the reconstructed lesion optical properties.

A variety of experimental and modeling approaches have been developed for system calibration of optical source strengths and detector gains, source and detector (optodes) position errors, and coupling errors between skin and optodes.³¹⁻³⁹ In general, source strengths and detector gains can be accurately estimated from homogeneous phantom measurements at each experiment and used to compensate tissue measurements. The compensation can be performed independently from image reconstruction^{31,33} or as part of the inverse problem for image reconstruction.^{32,34} These methods compensate system-related parameters and do not compensate tissue-caused wavelength-dependent measurement errors. Model-based approaches to compensate coupling errors between skin and optical optodes as well as optode position errors have been proposed in several studies.³⁵⁻³⁷ These methods model coupling errors as unknowns and include them in the reconstruction of tissue optical properties either sequentially or simultaneously. The advantage of these methods is the adaptive estimation of the coupling errors; however, the methods require the unknown coupling coefficients minimally vary from their initial estimates. Other approaches include the use of the differences between measurements at 2 separate wavelengths to reduce the coupling errors.³⁸ A nonlinear approach for difference imaging was studied on how this approach tolerates

modeling errors like domain truncation, source and detector coupling errors, and domain shape errors.³⁹ The challenges we are facing are wavelength-dependent measurement errors caused by low SNR and wavelength-dependent skin and optode coupling errors, which can vary depending on patients' bulk tissue optical properties, skin conditions, and operator's hand or patient motion during data acquisition. An effective and robust data processing method is needed to remove these measurement errors while preserving the wavelength-dependent lesion optical properties. The method also needs to be automated to minimize the user interface and facilitate clinical translation. In our early approach, the individual source strengths and detector gains were estimated using a least-square method and compensated before fitting each patient's background optical properties from the contralateral normal breast measurements.³¹ The perturbation approach, which is the difference between lesion and contralateral normal tissue measurements normalized to the contralateral measurements, was used in imaging reconstruction. This approach has canceled out the unknown source strengths and detector gains. In a recent attempt by our group, Vavadi and Zhu *et al*⁴⁰ have introduced a statistical method to automatically remove outliers from contralateral normal breast measurements. This method utilizes multiple sets of reference measurements to produce a robust set of reference. However, in many clinical cases, the outliers due to measurement errors are present in the lesion measurements. Lesion measurements contain wavelength-dependent information and are expected to be more heterogeneous than the reference measurements. To separate the measurement errors from lesion heterogeneity, more information from multiple wavelength measurements can be incorporated in the preprocessing before image reconstruction.

In this study, we introduce a new automated approach for data filtering based on multiple wavelength measurements collected at lesion site. The method combines data collected from multiple sets of lesion measurements to detect and correct outliers caused by wavelength-dependent measurement errors in the perturbation. This method represents an important step toward a fully automated DOT system for its clinical translation. The hypothesis is that there is a strong correlation between multiple wavelength measurements in NIR spectrum collected from the same lesion site and that this correlation has significantly decreased with the presence of wavelength-dependent measurement errors. The hypothesis was tested with phantom and patients' data.

Method

Ultrasound-Guided DOT System and Data Acquisition

Phantom experiments and clinical studies were performed using our US-guided DOT system. This system consists of 4 laser diodes of wavelength 740, 780, 808, and 830 nm and 10 parallel photomultiplier (PMT) detectors. Laser diodes were modulated at 140 MHz, and the light at each wavelength was sequentially delivered to 9 positions on a handheld probe through optical fibers. Ten light guides couple the reflected

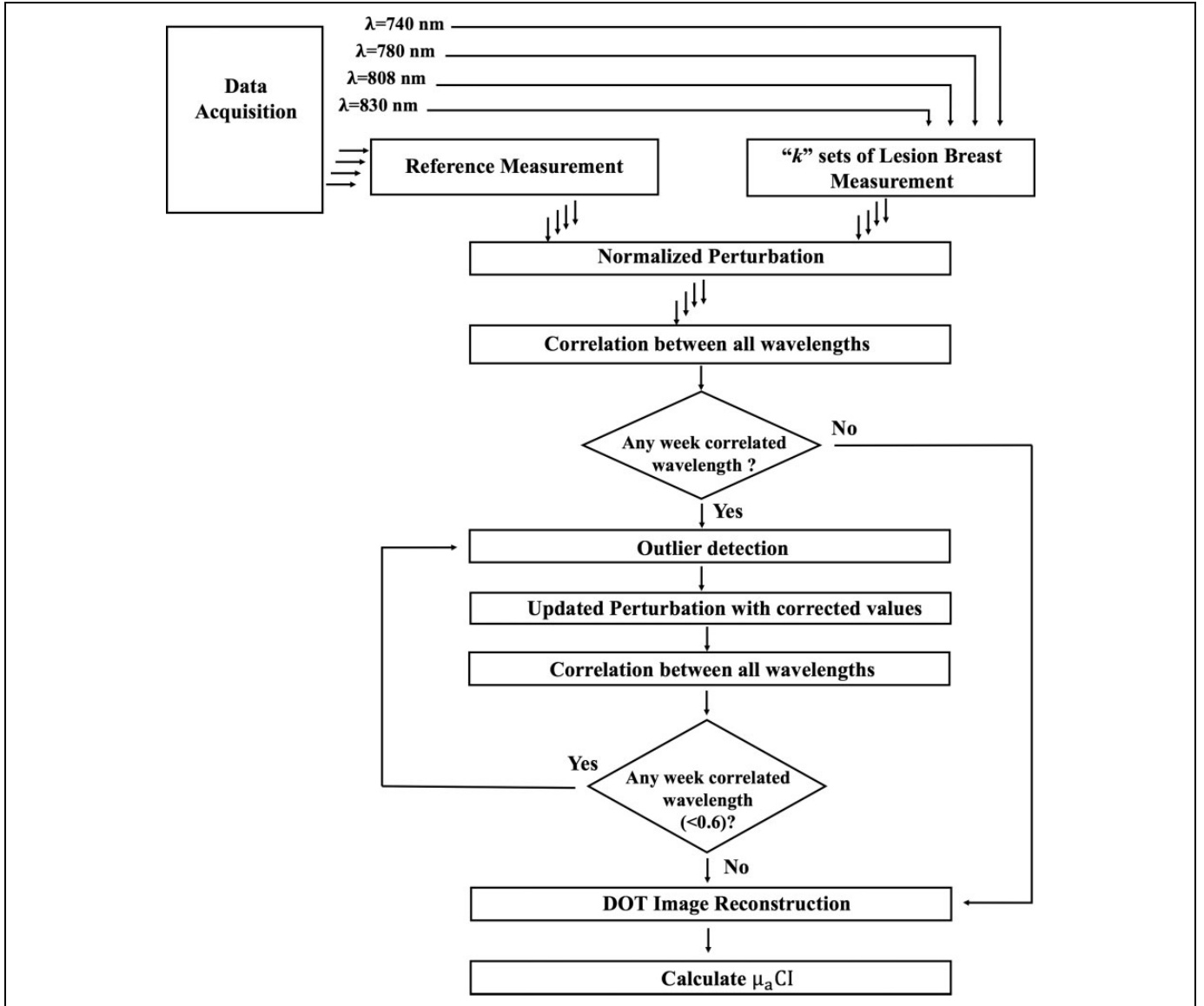


Figure 1. Block diagram of the proposed automated data preprocessing method.

light from tissue to 10 parallel PMT detectors simultaneously. As a result, each wavelength data set has 90 measurements. The details of the NIR system can be found elsewhere.¹⁹ To implement statistical tests for outlier detection, we repeated the data acquisition 3 times for each wavelength at the same lesion location, which results in 12 data sets of 90 measurements each.

Automated Preprocessing Approach

Using the US-guided DOT system, multiple sets of measurements are collected from a lesion site and a contralateral normal breast (a reference site) to calculate normalized perturbation. For each wavelength, the normalized perturbation (U_λ) is given as:

$$U_\lambda(m) = \frac{A_1 e^{j\varnothing_1(m)} - A_r e^{j\varnothing_r(m)}}{A_r e^{j\varnothing_r(m)}}, \quad (1)$$

where $A_1(m)$ and $\varnothing_1(m)$ are, respectively, measured amplitude and phase from the m th source–detector pair at lesion site, while $A_r(m)$ and $\varnothing_r(m)$, respectively, are the measured amplitude and phase from the m th source–detector pair at reference breast. This approach can cancel out the unknown source strengths and detector gains.³¹

As noted earlier, outliers in the normalized perturbation are mainly caused by measurement errors due to wavelength-dependent low SNR and skin–fiber coupling. In this method, perturbation outliers are detected and corrected by combining measurements of multiple wavelengths. A block diagram of the method is given in Figure 1. This figure illustrates that multiple data sets are acquired from a lesion site and a contralateral site of the normal breast. For each data set, our system is capable of acquiring measurements at 4 different wavelengths: 740, 780, 808, and 830 nm. The perturbation is calculated for each

wavelength measurement, which is the normalized difference between the examined lesion and the contralateral normal breast measurements. Next, the correlation coefficients between all wavelengths are calculated to find the weakest correlated wavelength measurement. An outlier detection method is applied to the weakest correlated wavelength measurement to detect the outliers and correct them. This process is followed until all outliers are corrected. The following text is a detailed explanation of the method.

In the first step, the normalized perturbations are used to calculate the correlation coefficients between each pair of wavelengths as^{41,42}:

$$\text{Corr}(\lambda_i, \lambda_j) = \frac{1}{M-1} \sum_{m=1}^M \left(\frac{U_{\lambda_i}(m) - \mu_{\lambda_i}}{\sigma_{\lambda_i}} \right) \left(\frac{U_{\lambda_j}(m) - \mu_{\lambda_j}}{\sigma_{\lambda_j}} \right), \quad (2)$$

where μ_{λ_i} and σ_{λ_i} are the mean and standard deviation of the normalized perturbations of the wavelength λ_i , respectively, and μ_{λ_j} and σ_{λ_j} are corresponding values of wavelength λ_j . M is the total number of measurements for each data set. The correlation coefficients are calculated for both the real and imaginary parts of perturbation. Based on the average correlation coefficients of both real and imaginary parts, the measurement set with the weakest correlated wavelength with correlation value below 0.6 determined from phantom experiments is selected for outlier detection. If all correlation coefficients are above 0.6, the reconstruction will be performed for each wavelength and then total hemoglobin concentration (tHb) will be calculated.

In the second step, the method examines each source–detector pair of both real and imaginary perturbation in the selected wavelength to determine outliers. This is achieved by combining 4 wavelength data of 3 lesion data sets, so that for each source–detector pair we have 12 measurements. Using multiple data sets measured at each source–detector pair, we are able to apply statistical test to detect outlier measurements for each source–detector pair and eliminate them.

We use the maximum normed residual (MNR) method for outlier detection. The MNR is a statistical test used to detect outliers based on the largest absolute deviation from the sample mean. By calculating the t distribution with $k-2$ degrees of freedom, a threshold value for each source–detector pair i can be obtained as:

$$G_{\text{Threshold}}(i) = \frac{k-1}{\sqrt{k}} \sqrt{\frac{t_{\frac{\alpha}{2}, k-2}^2(i)}{k-2 + t_{\frac{\alpha}{2}, k-2}^2(i)}}, \quad (3)$$

where $G_{\text{Threshold}}(i)$ is the outlier threshold for i th source–detector pair, $t_{\frac{\alpha}{2}, k-2}^2(i)$ denotes the upper critical value of the t distribution with $k-2$ degrees of freedom, and α represents the level of significance that determines the strictness of outlier removal procedure. By changing this value between 0 and 1, the total number of the outliers and the significance of

these outliers removed from the database can be changed. To find the optimal value of α , the outlier removal process is performed for different significance level ranging from .01 to .5 and the optimal value is set to .05 based on visual examination of the removed outliers. This optimal value is selected in a way that the test only removes the significant outlier data. A G value is determined as an absolute deviation of the data point from mean value of the measurements and normalized by standard deviation. The data point corresponding to the maximum G value which has absolute deviation higher than the threshold is considered as an outlier and removed from the data set. The test is iterated until no further outliers are detected beyond the threshold. This test is done for real and imaginary parts of perturbations separately. The details of this method can be found elsewhere.^{40,42-44} If an outlier is detected for any source–detector pair, its value is corrected with the average value of other wavelengths of the same source–detector pair. Both the real and imaginary parts are updated.

Once outliers are corrected for the selected wavelength, the correlation coefficients are recalculated between wavelengths with the updated values. If there are still any other weakly correlated wavelengths measurements (correlation < 0.6), the above procedure is iteratively followed until all wavelengths measurements are corrected. The corrected normalized perturbation is used for DOT image reconstruction.

Diffuse Optical Tomography Image Reconstruction

The US-guided dual-zone mesh method introduced by our group^{9,19} was utilized to perform the DOT image reconstruction. In this framework, we use the coregistered US image to divide the DOT imaging volume into a lesion region as a region of interest and a background region. A finer mesh is selected for the lesion and a coarse mesh for the background region. Thus, the method employs smaller voxel size for lesion and larger coarse voxel size for the background region. This technique notably reduces the total number of voxels with unknown optical properties and keeps it at the same scale as of the total measurements. As a result, the ill-posed DOT reconstruction problem has significantly improved. Born approximation of light propagation in the tissue is utilized for computing the weight matrix and an iterative optimization based on conjugate gradient method is used to compute the lesion distribution at each wavelength.

The reconstructed absorption maps of the 4 wavelengths (740, 780, 808, and 830 nm) are used to compute the tHb map directly using the summation of oxygenated hemoglobin and deoxygenated hemoglobin maps with extinction coefficients given in reference.⁴⁵

Computation of Wavelength Consistency Index

To assess the improvement in consistency between reconstructed results obtained from different wavelength

measurements, we calculate the absorption coefficient μ_a consistency index (μ_a CI) as:

$$\mu_a \text{CI} = \frac{1}{V} \sum_{p=1}^V \left\{ \frac{1}{W(W-1)/2} \left[\sum_{i=1}^W \sum_{j=i+1}^W |\mu_a^{\lambda_i}(p) - \mu_a^{\lambda_j}(p)| \right] \right\}, \quad (4)$$

where “V” is number of voxels and “W” is number of wavelengths. This index is calculated voxel by voxel for absorption coefficient $\mu_a(p)$ of each pair of wavelengths. The value of this index is used as an indicator to compare the improvement in the consistency between wavelengths before and after the application of the method. Since this index calculates the absolute difference between absorption maps at different wavelengths, a low index value is an indicator of high absorption consistency between wavelengths. Likewise, a high index value is an indicator of low absorption consistency between wavelengths.

To evaluate the range of μ_a CIs obtained from breast lesions without evidence of wavelength-dependent measurement errors, we have applied the correction method to 12 patients’ data. For each patient’s data, the correlation coefficient between each pair of wavelength measurements was calculated and found to be larger than 0.6, and μ_a CIs before and after correction were calculated. Result was used as a reference to characterize breast lesion wavelength-dependent absorption heterogeneity.

Results

Phantom Study

We performed phantom study to evaluate the correlation coefficients between measurements of multiple wavelengths at the same target site. In phantom experiments, we used Intralipid solution as a reference to calculate the normalized perturbation. The Intralipid solution is a homogenous medium, which eliminates any reference heterogeneity effect on the calculated perturbation. Moreover, the DOT handheld probe was fixed during data acquisition in the phantom experiments. Outliers caused by patient’s or user’s movements of the handheld DOT probe during data acquisition did not exist in the phantom data.

A phantom target with calibrated value of $\mu_a = 0.23 \text{ cm}^{-1}$ and diameter 3 cm was used for experiments. The target was merged in the Intralipid solution (which calibrated as $\mu_a = .03 \text{ cm}^{-1}$) at different depths of 2.0 to 3.5 cm in 0.5 cm increments. The same Intralipid solution was used to acquire the reference measurements. Table 1 shows the correlation coefficients between measurements of multiple wavelengths for the target. Here, the presented correlation coefficients are the average correlation values of the real and imaginary parts of the perturbation. The table also shows the reconstructed target absorption coefficients at different depths. The phantom experiments show strong correlation coefficients between measurements of different wavelengths. The correlation coefficients decrease with depths due to lower SNR of the

Table 1. The Correlation Coefficients, Maximum Reconstructed Absorption, and μ_a CI Values for the Phantom Target at Different Depths.

	Depth (cm)			
	2	2.5	3	3.5
Correlation coefficients				
Corr (740, 780)	0.9561	0.9223	0.7706	0.7130
Corr (740, 808)	0.9004	0.9564	0.9545	0.7063
Corr (740, 830)	0.9884	0.9728	0.9699	0.8797
Corr (780, 808)	0.9494	0.9119	0.7803	0.7409
Corr (780, 830)	0.9252	0.9738	0.8411	0.7711
Corr (808, 830)	0.9798	0.9660	0.9654	0.8720
Maximum reconstructed μ_a (cm^{-1})				
$\lambda = 740 \text{ nm}$	0.1093	0.1744	0.2435	0.2230
$\lambda = 780 \text{ nm}$	0.0931	0.1565	0.2280	0.1994
$\lambda = 808 \text{ nm}$	0.0832	0.1628	0.2363	0.2124
$\lambda = 830 \text{ nm}$	0.0894	0.1767	0.2548	0.2283
μ_a CI	0.0137	0.0159	0.0269	0.0476

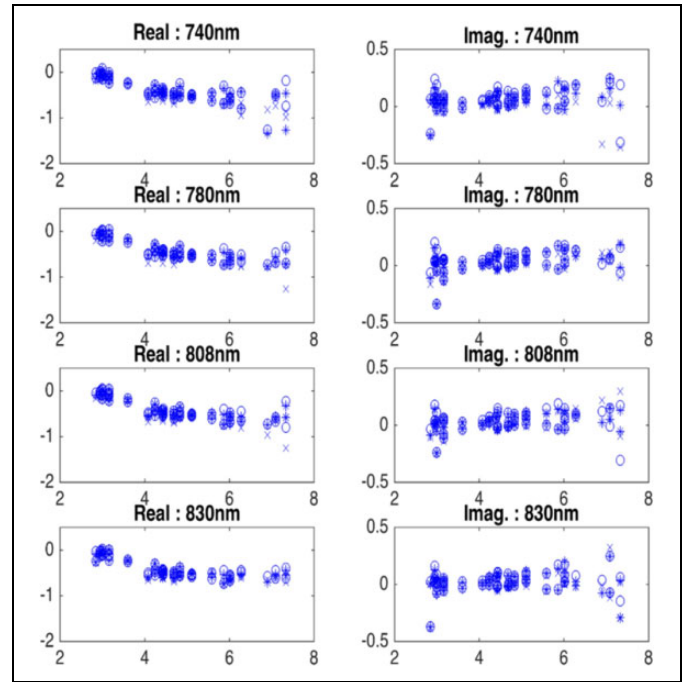


Figure 2. Real and imaginary plots of normalized perturbation (equation 1) of a wavelength consistent case with no data preprocessing. x-Axis is the source and detector distance in centimeter.

measurements, but all coefficients are above 0.6. Also, we see that reconstructed absorption coefficients are consistent among all wavelengths (Table 1). However, consistency between wavelengths also decreases with depth; the μ_a CI does not exceed 0.06 in the reconstructed absorption profiles obtained at different depth. Note that the lower sensitivity at the shallower depth 2 cm (target center depth) is related to the lack of shorter source and detector pairs across the center of the probe where the US transducer is housed. This low sensitivity can be calibrated.

Table 2. The Correlation Coefficients, Maximum Reconstructed Absorption, and μ_a CI of a Wavelength Consistent Case.

	No Preprocessing
Correlation coefficients	
Corr(740, 780)	0.91
Corr (740, 808)	0.89
Corr (740, 830)	0.79
Corr (780, 808)	0.92
Corr (780, 830)	0.89
Corr (808, 830)	0.84
Maximum reconstructed μ_a (cm^{-1})	
$\lambda = 740$ nm	0.225
$\lambda = 780$ nm	0.216
$\lambda = 808$ nm	0.237
$\lambda = 830$ nm	0.226
μ_a CI	0.023

Clinical Study

In this study, we have examined a total of 28 patients. These patients are divided into 2 groups based on the wavelength correlation analysis and μ_a CI values. Sixteen patients' data were wavelength inconsistent cases, where the wavelength correlation coefficients were low and μ_a CIs were high. Based on biopsy results, 8 of these patients had malignant lesions and 8 had benign lesions. The other group includes 12 wavelength consistent cases, where the wavelength correlation coefficients were high and μ_a CI was low. The study was approved by institutional review boards and all patients signed the informed consent. The data used for this study were deidentified.

In Figure 2, we present the real and imaginary plots versus source–detector distance of normalized perturbation profiles for all 4 wavelengths of a wavelength consistent case with no

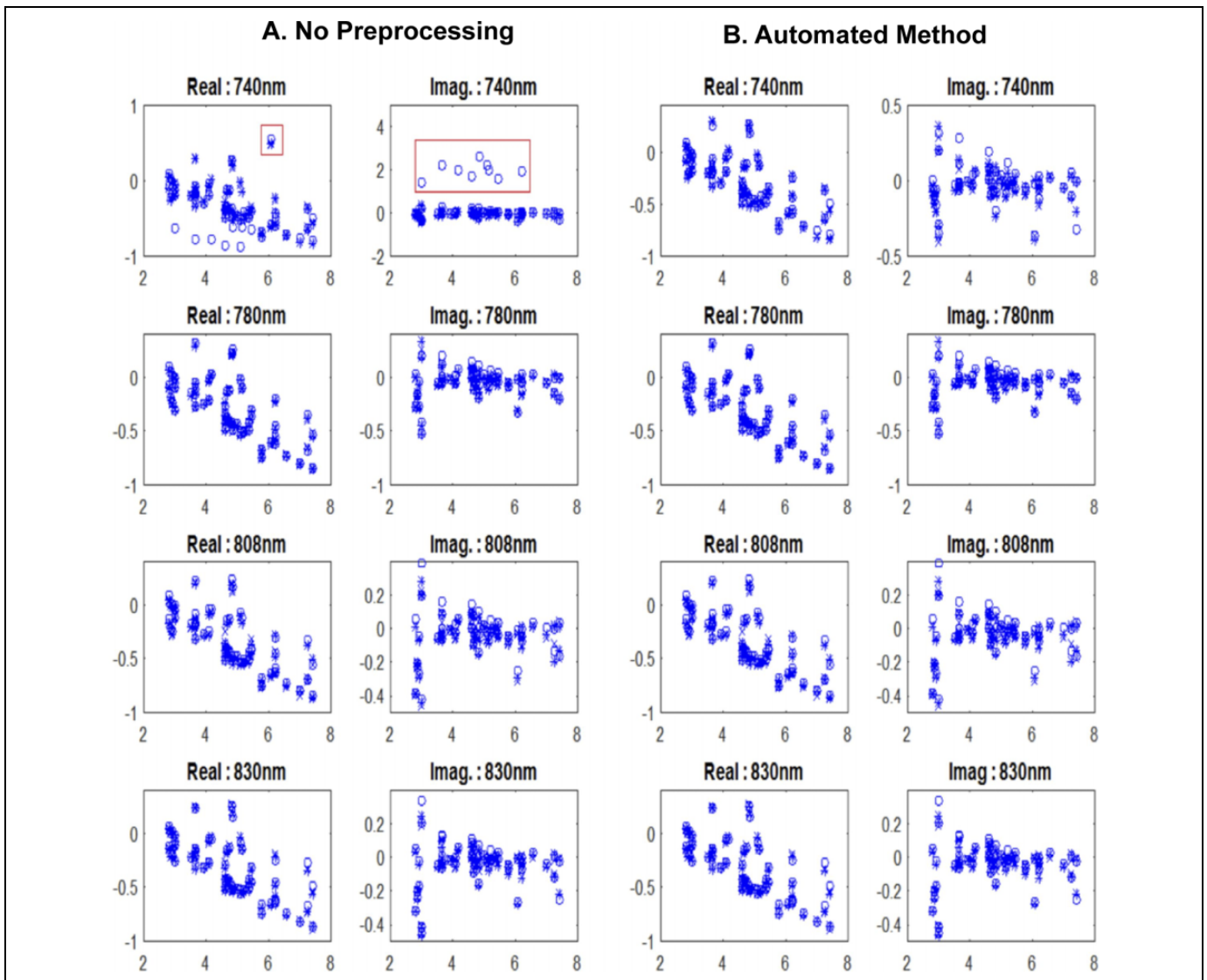


Figure 3. Real and imaginary plots of normalized perturbation (equation 1) of a malignant lesion with (A) no data preprocessing and (B) after applying the proposed automated method. x-Axis is the source and detector distance in centimeter.

Table 3. The Correlation Coefficients, Maximum Reconstructed Absorption, and μ_a CI Values Before and After Automated Method for a Malignant Case.

	No Preprocessing	Automated Preprocessing Method
Correlation coefficients		
Corr (740, 780)	0.50	0.81
Corr (740, 808)	0.39	0.75
Corr (740, 830)	0.79	0.82
Corr (780, 808)	0.91	0.91
Corr (780, 830)	0.74	0.74
Corr (808, 830)	0.73	0.73
Maximum reconstructed μ_a (cm^{-1})		
$\lambda = 740$ nm	0.52	0.33
$\lambda = 780$ nm	0.36	0.36
$\lambda = 808$ nm	0.33	0.33
$\lambda = 830$ nm	0.35	0.35
μ_a CI	0.10	0.04

preprocessing. Both real and imaginary parts of each wavelength show similar trend with other wavelengths, with no visible outliers. The range of the real and imaginary parts of the perturbation is similar across 4 wavelengths.

Table 2 shows the correlation coefficients, the maximum reconstructed absorption coefficients of all wavelengths, and μ_a CI for the wavelength consistent case presented in Figure 2. In this case, we see higher correlation between all wavelengths and the maximum reconstructed absorption coefficients of all wavelengths are in the same range. In addition, μ_a CI shows low value, which indicates high consistency in μ_a between wavelengths.

We have applied the automated preprocessing method to all cases. For wavelength inconsistent cases, measurements with outliers are seen to have weak correlation coefficients with others. An example of a normalized perturbation of a malignant case is illustrated in Figure 3. This figure shows the calculated perturbation profile for each wavelength versus source–detector distance. The wavelength 740 nm is weakly correlated with the other wavelengths, with correlation coefficients of less than 0.6, as summarized in Table 3. This lower correlation value is due to outliers observed in both real and imaginary parts of the 740 nm wavelength shown in Figure 3A. When the automated method is applied, outliers are corrected (Figure 3B) for the 740 nm wavelength and the correlation coefficients are improved (Table 3).

The reconstructed absorption maps of this case are illustrated in Figure 4. Both the reconstructed lesion shape and the maximum reconstructed absorption value for the wavelength 740 nm are not consistent with the other wavelengths (Figure 4A and Table 3). After applying the automated method, we see the shape of this wavelength becomes more consistent with other wavelengths (Figure 4B), and μ_a CI reduced from .10 to .04. Likewise, the correlation coefficients between all wavelength measurements are improved (Table 3).

A benign case is also presented in Figure 5. The wavelength correlation coefficients and the reconstructed absorption coefficients are summarized in Table 4. Based on the wavelength correlation analysis, the wavelength 808 nm is weakly correlated with the other wavelengths. Table 4 summarizes the average correlation coefficients of real and imaginary parts of the perturbation between multiple wavelengths. With no perturbation preprocessing, the reconstructed absorption map at 808 nm is too high and distorted as compared to other wavelengths. When the automated method is used, the μ_a CI has decreased from 0.14 to 0.05. The maximum reconstructed absorption coefficient of the wavelength 808 nm became more consistent with the rest of the wavelengths and decreases from 0.27 to 0.13 cm^{-1} . Besides that, the tHb decreased from 68 to 55 $\mu\text{mol/L}$ when the automated method was used.

An example of 2 inconsistent wavelength measurements is presented in Table 5. In this example, the wavelengths 740 and 830 nm are weakly correlated with 2 other wavelengths (low correlation coefficients). The correlation between the wavelengths 780 and 808 nm is high. With no preprocessing, the reconstructed absorption coefficient is too high at 740 nm and too low at 830 nm. We applied the automated method to this case and correlation coefficients have improved for the wavelengths 740 and 830 nm. The maximum reconstructed absorption coefficient for 740 nm decreased from 0.35 to 0.16, and the maximum reconstructed absorption for 830 nm has increased to 0.14 and became more consistent with the rest of the wavelengths. Also, the wavelength consistency index (μ_a CI) reduced from 0.12 to 0.03.

The average value of the μ_a CI for all wavelength inconsistent clinical cases has reduced from 0.12 to 0.04 when the automated method is applied, as shown in Figure 6. For the wavelength consistent clinical cases, the average μ_a CI with no preprocessing is 0.03 and has remained the same after applying the automated method (Figure 6). This indicates an improvement in the consistency of the reconstructed absorption coefficients between all wavelengths for the wavelength inconsistent cases.

The average wavelength correlation coefficients calculated for all 28 clinical cases are summarized in Table 6. Here, the wavelength correlation coefficients of the wavelength inconsistent cases are compared to those of the wavelength consistent cases. The results show the correlation coefficients of the wavelength inconsistent cases after applying the automated method are similar to the correlation coefficients of the wavelength consistent cases.

The reconstructed absorption coefficients of the 4 wavelengths (740, 780, 808, and 830 nm) are used to compute the tHb. The computed tHb for the wavelength inconsistent cases is presented in Figure 7, which shows the calculated tHb for malignant and benign cases before and after applying the automated method. A 2-sided *t* test was performed between malignant and benign groups before and after applying the automated method. The proposed method shows the statistical significance, where the *P* value improved from .006 to .001 using proposed automated method. The tHb ratio of malignant

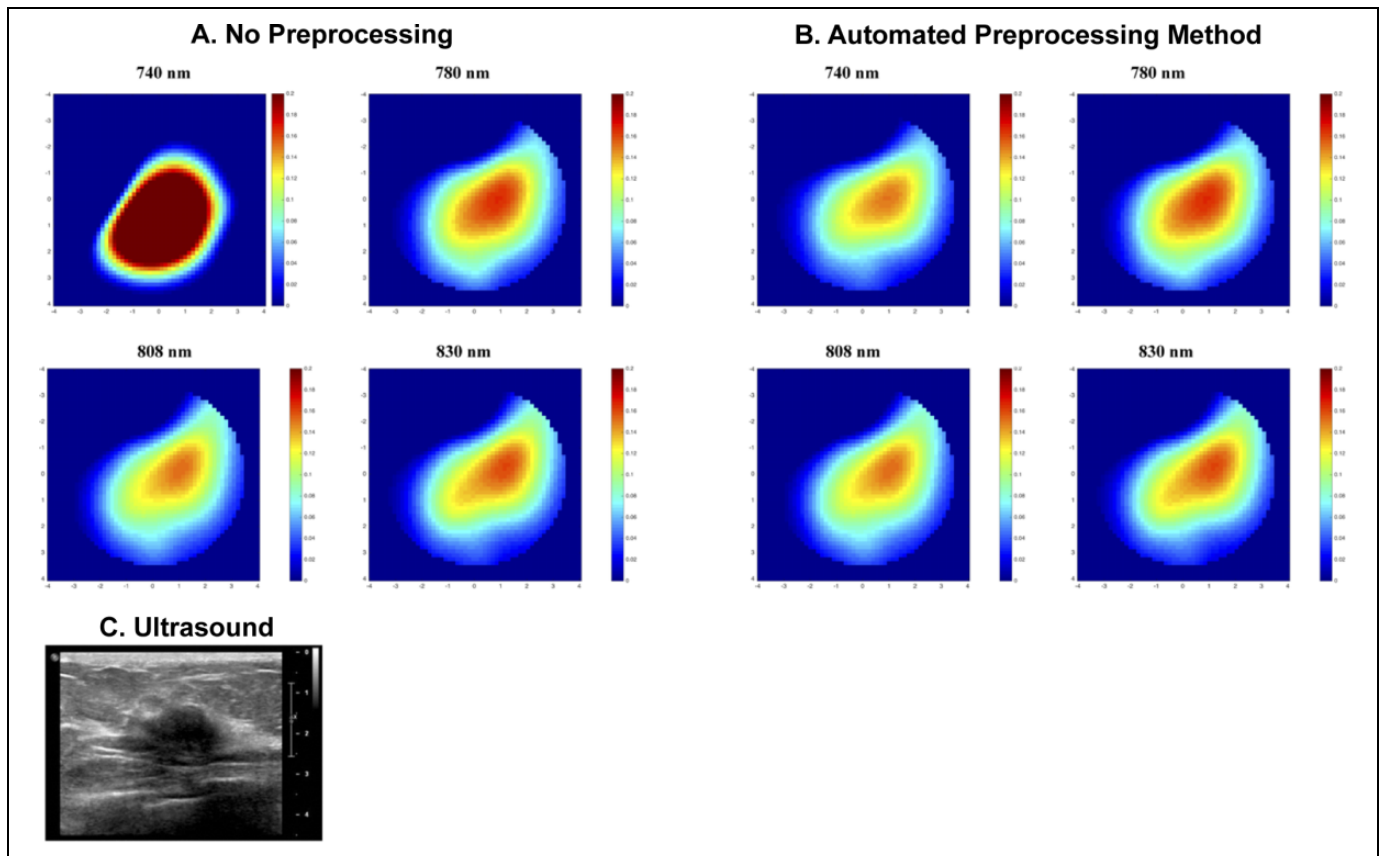


Figure 4. The reconstructed absorption maps of a malignant lesion (A and B) along with the ultrasound B scan of the lesion (C). Reconstructed absorption coefficients are shown for 4 wavelengths with (A) no data preprocessing and (B) proposed automated method. Each image marked with wavelength represents a spatial image in x and y dimensions of 9 cm scale at the target depth of 2 cm, other depths are not shown.

to benign group is improved from 2.01 to 2.42 with the automated preprocessing method.

Discussion and Summary

We have introduced a new automated data preprocessing method based on multiple wavelength measurements. In the absence of wavelength-dependent measurement errors in phantom experiments, the data show strong correlations between measurements of multiple wavelengths in the NIR range. In addition, the phantom experiments show a high consistency between reconstructed images obtained from measurements of different wavelengths. Even though both the correlation and the consistency between wavelengths are reduced with target depth, correlation coefficients have maintained beyond 0.6. Similarly, the calculated μ_a CI has not exceeded .06 for all target depths in the phantom experiments.

The automated preprocessing method was used to analyze 16 malignant and benign wavelength inconsistent cases. The results of this method were compared with those obtained with no data preprocessing. The wavelength correlation coefficients and the consistency indexes, μ_a CI, were further compared with those values obtained from 12 wavelength consistent cases. Both the correlation coefficients and

the μ_a CI of the wavelength consistent cases and the wavelength inconsistent cases after applying the automated method are in the same range (Table 6). Because the correlation coefficients for wavelength consistent cases were above 0.6 and correlation coefficients at all depths in the phantom study were also above 0.6, we have chosen 0.6 as a threshold value in this method. The automated method helped to improve the consistency between reconstructed absorption maps of different wavelengths for all malignant and benign cases (Figures 4 and 5). The high μ_a CI values of wavelength inconsistent cases have been improved after applying the automated preprocessing method. These values are all decreased to the similar values of wavelength consistent cases and to phantom data μ_a CI values of less than .06. This improvement in the μ_a CI shows that the wavelength-dependent measurement errors have minimized using other wavelength measurements to improve the data correlation before performing reconstruction.

The lesion shape is also improved and become more consistent with the results of other wavelengths as we see in Figures 4 and 5. With no preprocessing, the reconstructed absorption coefficients of the benign case (Figure 5) at 808 nm are high and distorted as compared to the other wavelengths. The automated method demonstrated the ability to

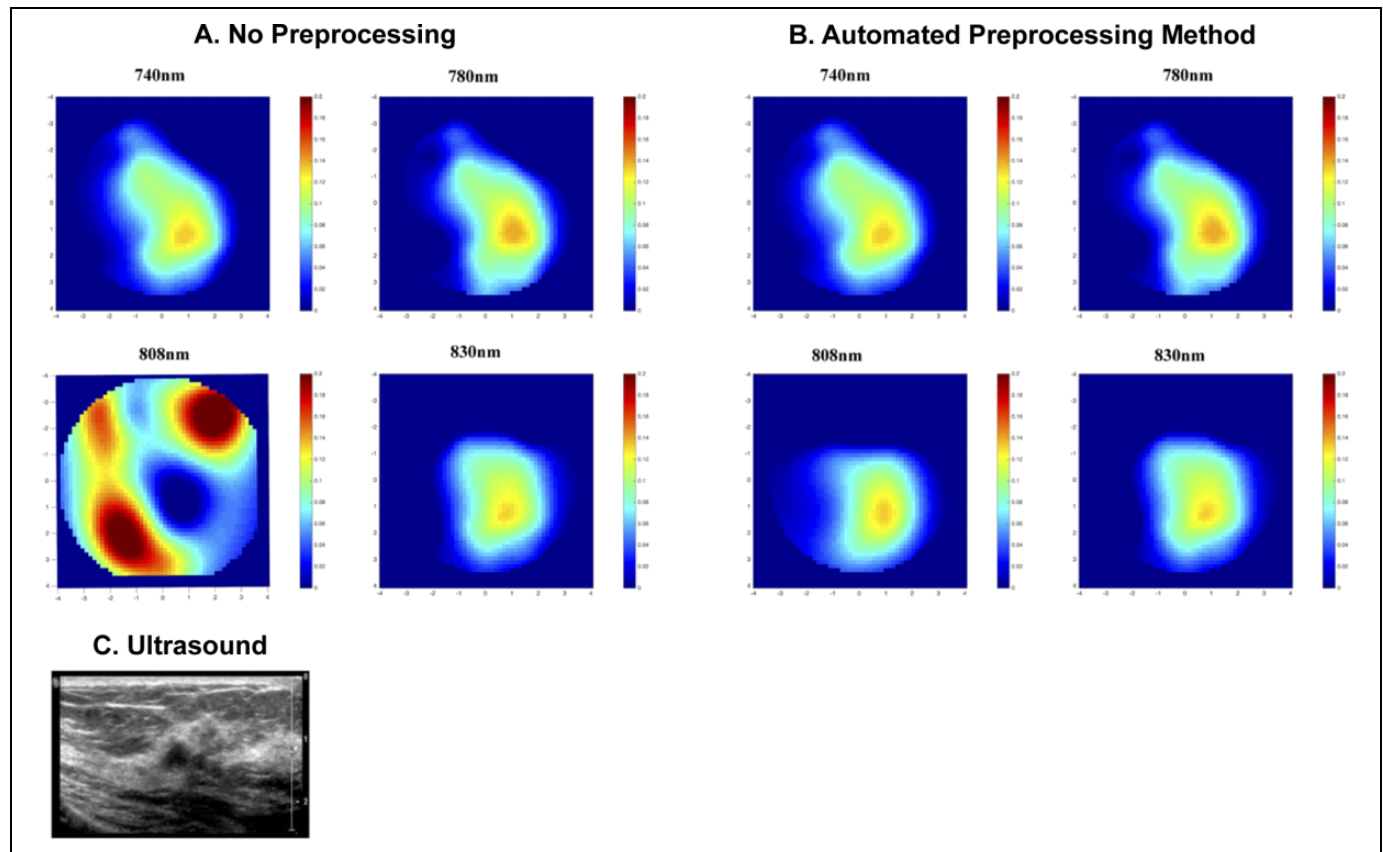


Figure 5. The reconstructed absorption map of a benign lesion (A and B) along with the ultrasound B scan of the lesion (C). Reconstructed absorption coefficients are shown for 4 wavelengths with (A) no data preprocessing and (B) proposed automated method.

Table 4. The Correlation Coefficients, Maximum Reconstructed Absorption, and μ_a CI Values Before and After Automated Method for a Benign Case.

	No Preprocessing	Automated Preprocessing Method
Correlation coefficients		
Corr (740, 780)	0.72	0.72
Corr (740, 808)	0.33	0.77
Corr (740, 830)	0.76	0.76
Corr (780, 808)	0.19	0.67
Corr (780, 830)	0.79	0.79
Corr (808, 830)	0.20	0.81
Reconstructed μ_a (cm^{-1})		
$\lambda = 740$ nm	0.148	0.148
$\lambda = 780$ nm	0.154	0.154
$\lambda = 808$ nm	0.276	0.137
$\lambda = 830$ nm	0.152	0.152
μ_a CI	0.14	0.05

Table 5. The Correlation Coefficients, Maximum Reconstructed Absorption, and μ_a CI Values Before and After Automated Method for a Case With 2 Inconsistent Wavelengths.

	No Preprocessing	Automated Preprocessing Method
Correlation coefficients		
Corr (740, 780)	0.21	0.74
Corr (740, 808)	0.35	0.79
Corr (740, 830)	0.36	0.71
Corr (780, 808)	0.87	0.87
Corr (780, 830)	0.47	0.85
Corr (808, 830)	0.45	0.89
Reconstructed μ_a (cm^{-1})		
$\lambda = 740$ nm	0.354	0.167
$\lambda = 780$ nm	0.155	0.155
$\lambda = 808$ nm	0.156	0.156
$\lambda = 830$ nm	0.125	0.141
μ_a CI	0.12	0.03

improve the consistency for this case. The shape of the lesion at 808 nm wavelength became more consistent as well.

These improvements were reflected in the tHb. The average value of the tHb for malignant cases increased by 10 $\mu\text{mol/L}$, while benign group showed similar improvement with tHb reduced by 12 $\mu\text{mol/L}$. As a result, the malignant to

benign tHb ratio was improved by 20% after applying the automated method.

The automated method used in this study focuses only on the cases with one poorly correlated wavelength with other wavelengths, and all analyzed inconsistent cases are having problems in one wavelength. The results show strong

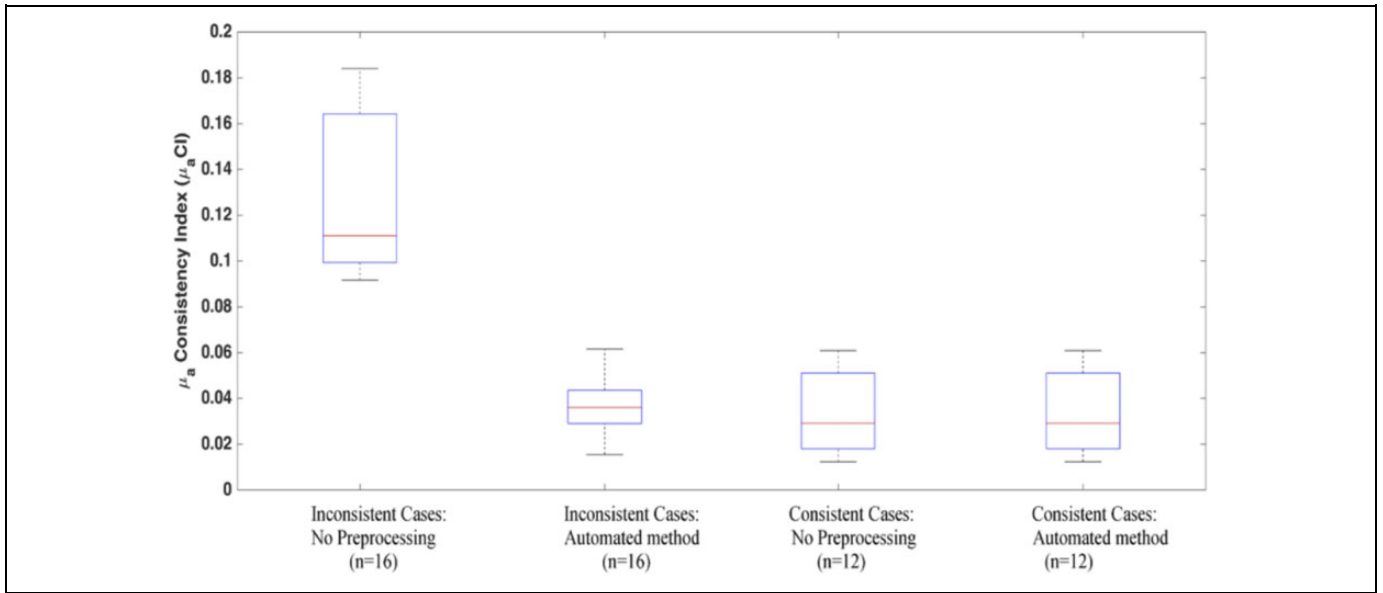


Figure 6. The calculated μ_a CI values for the 2 groups of clinical cases: the wavelength inconsistent cases and the wavelength consistent cases. The figure shows the index values of each group before and after the automated method is used.

Table 6. Average (Standard Deviation) of the Correlation Coefficients of the Clinical Cases.

	Wavelength Inconsistent Cases: No Preprocessing	Wavelength Inconsistent Cases: Automated Method	Wavelength Consistent Cases: No Preprocessing
Wavelength correlation coefficients	0.35 (0.22)	0.89 (0.09)	0.83 (0.22)

improvement in terms of improving the wavelength consistency and robustness of the image reconstruction method. In cases that more than one wavelength measurements are distorted, the method can still be applied using the remaining consistent wavelength measurements to predict the distorted wavelengths and correct the distortion. However, our method requires that the wavelength with measurement outliers should be at most 2 if 4 optical wavelengths are used.

There are 4 major steps for a fully automated US-guided DOT system. First, the system calibration should be done automatically and the code has been implemented in our imaging algorithm. Second, the estimation of background optical properties of the contralateral breast and therefore the computation of the weight matrix should be done automatically after filtering out outliers in the measurements. Third, the perturbation, which is the normalized difference between the examined lesion and the background, should be automated after correcting wavelength-dependent errors and also eliminating other noise due to low SNR. Fourth, the US lesion measurements should be done automatically and these parameters can be inputted into the DOT imaging reconstruction for near-real-time US-guided DOT imaging. In this study, the automated approach for data filtering based on multiple wavelength measurements collected at lesion site is intended to automate the third step data processing, while the method introduced in Vavadi and Zhu⁴⁰ was used to automate the second step data processing. The semiautomated method proposed by Mostafa *et al*²⁵ is aimed to solve problems of the fourth step. The future work will combine these 4 steps of data preprocessing and validate and refine the combined approach in a large prospective patient study.

In conclusion, this study has introduced an automated data preprocessing method based on measurements of multiple wavelengths. The method eliminates the effect of

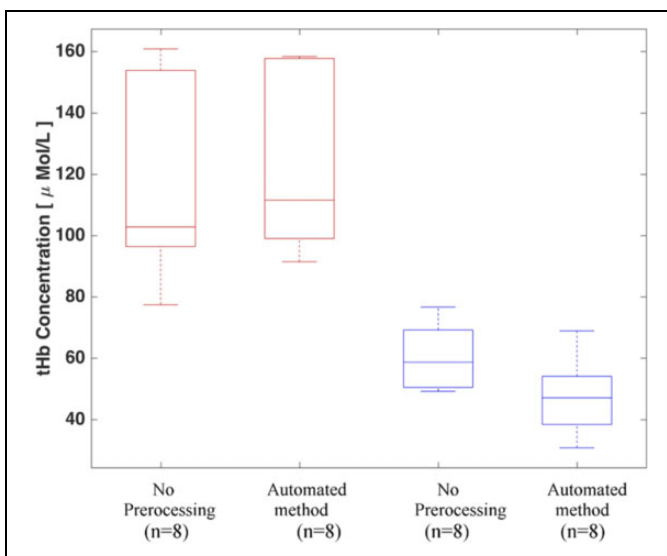


Figure 7. Total hemoglobin concentration of wavelength inconsistent clinical cases before and after the automated data preprocessing method. Red and blue boxes indicate the malignant and benign cases, respectively.

wavelength-dependent measurement errors in the DOT perturbation, which helps to achieve more accurate reconstruction of optical properties of breast lesions. The automated method also helps to minimize both the user interface and the time for data preprocessing. The average time for the experienced user to manually perform data preprocessing is from 15 to 30 minutes for one patient's data. The automated method could help to reduce this time to less than a minute. Although the method is demonstrated using US-guided DOT data, it is applicable to any DOT data preprocessing obtained with multiple wavelengths.

Authors' Note

This study used patients' data but de-identified. IRB #02-064S-2 (2002-2015).


Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors acknowledge and thank the funding supports of this work from NIH (EB002136), Connecticut Bioscience Innovation Fund (CBIF) Award #513. Murad Althobaiti acknowledges the funding support of fellowship from Saudi Arabian Cultural Mission of the Royal Embassy of Saudi Arabia.

ORCID iD

Hamed Vavadi, PhD  <http://orcid.org/0000-0002-3054-0169>

Quing Zhu, PhD  <http://orcid.org/0000-0002-1837-2588>

References

- Gibson AP, Hebden JC, Arridge SR. Recent advances in diffuse optical imaging. *Phys Med Biol*. 2005;50(4):R1-R43. <http://eprints.ucl.ac.uk/2998/%5Cnhttp://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=15773619&retmode=ref&cmd=prlinks%5Cnhttp://iopscience.iop.org/0031-9155/50/4/R01>.
- Taroni P, Pifferi A, Torricelli A, Comelli D, Cubeddu R. In vivo absorption and scattering spectroscopy of biological tissues. *Photochem Photobiol Sci*. 2003;2(2):124.
- Chance B, Nioka S, Zhang J, et al. Breast cancer detection based on incremental biochemical and physiological properties of breast cancers: a six-year, two-site study. *Acad Radiol*. 2005;12(8):925-933.
- Quarto G, Spinelli L, Pifferi A, et al. Estimate of tissue composition in malignant and benign breast lesions by time-domain optical mammography. *Biomed Opt Express*. 2014;5(10):3684-3698. <http://www.osapublishing.org/boe/abstract.cfm?URI=boe-5-10-3684>.
- Boas DA, Brooks DH, Miller EL, et al. Imaging the body with diffuse optical tomography. *IEEE Signal Process Mag*. 2001;18(6):57-75.
- EGgebrecht AT, Ferradal SL, Robichaux-Viehoever A, et al. Mapping distributed brain function and networks with diffuse optical tomography. *Nat Photonics*. 2014;8(6):448-454. doi:10.1038/nphoton.2014.107.
- Owell P, Uaggia SIQ, Ighton DAH, et al. Functional imaging of the human brain using a modular, fibre-less, high-density diffuse optical tomography system. *Biomed Opt Express*. 2016;7(10):4275-4288.
- Durduran T, Choe R, Baker WB, Yodh AG. Diffuse optics for tissue monitoring and tomography. *Rep Prog Phys*. 2010;73(7):76701. doi:org/0034-4885/73/i=7/a=076701.
- Zhu Q, Chen N, Kurtzman SH. Imaging tumor angiogenesis by use of combined near-infrared diffusive light and ultrasound. *Opt Lett*. 2003;28(5):337-339. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1533768/>.
- Culver JP, White BR, Schlaggar BL, Dehghani H, Zeff BW. Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. *PNAS*. 2007;104(29):12169-12174. <http://www.ncbi.nlm.nih.gov/pubmed/17616584>.
- Chen H, Lin Z, Mo L, Wu T, Tan C. Near-infrared spectroscopy as a diagnostic tool for distinguishing between normal and malignant colorectal tissues. *Biomed Res Int*. 2015;2015:7.
- Zhang Q, Brukilacchio TJ, Li A, et al. Coregistered tomographic x-ray and optical breast imaging: initial results. *J Biomed Opt*. 2005;10(2):024033. doi:10.1117/1.1899183.
- Brooksby B, Pogue BW, Jiang S, et al. Imaging breast adipose and fibroglandular tissue molecular signatures by using hybrid MRI-guided near-infrared spectral tomography. *Proc Natl Acad Sci U S A*. 2006;103(23):8828-8833. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1482663/>.
- Fang Q, Selb J, Carp SA, et al. Combined optical and x-ray tomosynthesis breast imaging. *Radiology*. 2011;258(1):89-97. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3009384/>.
- Ntziachristos V, Yodh AG, Schnall MD, Chance B. MRI-guided diffuse optical spectroscopy of malignant and benign breast lesions. *Neoplasia*. 2002;4(4):347-354. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1661680/>.
- Davis SC, Dehghani H, Wang J, Jiang S, Pogue BW, Paulsen KD. Image-guided diffuse optical fluorescence tomography implemented with Laplacian-type regularization. *Opt Express*. 2007;15(7):4066-4082. <http://www.ncbi.nlm.nih.gov/pubmed/19532650>.
- Güven M, Yazıcı B, İntes X, Chance B. Diffuse optical tomography with a priori anatomical information. *Phys Med Biol*. 2005;50(12):2837-2858. <http://iopscience.iop.org/article/10.1088/0031-9155/50/12/008>.
- Zhang L, Zhao Y, Jiang S, Pogue BW, Paulsen KD. Direct regularization from co-registered anatomical images for MRI-guided near-infrared spectral tomographic image reconstruction. *Biomed Opt Express*. 2015;6(9):3618-3630. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4574684&tool=pmcentrez&rendertype=abstract>.
- Zhu Q, Xu C, Guo P, et al. Optimal probing of optical contrast of breast lesions of different size located at different depths by us localization. *Technol Cancer Res Treat*. 2006;5(4):365-380. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2018682/>.
- Althobaiti M, Vavadi H, Zhu Q. Diffuse optical tomography reconstruction method using ultrasound images as prior for

- regularization matrix. *J Biomed Opt.* 2017;22(2):26002. <http://biomedicaloptics.spiedigitallibrary.org/article.aspx?doi:10.1117/1.JBO.22.2.026002>.
21. Zhu Q, Tannenbaum S, Hegde P, Kane M, Xu C, Kurtzman SH. Noninvasive monitoring of breast cancer during neoadjuvant chemotherapy using optical tomography with ultrasound localization. *Neoplasia.* 2008;10(10):1028-1040.
 22. Zhu Q, Hegde PU, Ricci A, et al. Early-stage invasive breast cancers: potential role of optical tomography with us localization in assisting diagnosis. *Radiology.* 2010;256(2):367-3678. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2909434/>.
 23. Xu C, Vavadi H, Merkulov A, et al. Ultrasound-guided diffuse optical tomography for predicting and monitoring neoadjuvant chemotherapy of breast cancers: recent progress. *Ultrason Imaging.* 2016;38(1):5-18. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5056904/>.
 24. Zacharopoulos A, Schweiger M, Kolehmainen V, Arridge S. 3D shape based reconstruction of experimental data in diffuse optical tomography. *Opt Express.* 2009;17(21):18940-18956.
 25. Mostafa A, Vavadi H, Uddin KMS, Zhu Q. Diffuse optical tomography using semiautomated coregistered ultrasound measurements. *J Biomed Opt.* 2017;22(12):1-12. doi:10.1117/1.JBO.22.12.121610.
 26. Zhu Q, Huang M, Chen N, et al. Ultrasound-guided optical tomographic imaging of malignant and benign breast lesions: initial clinical results of 19 cases. *Neoplasia.* 2003;5(5):379-388. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1502608/>.
 27. Zhu Q, Ricci A, Hegde P, et al. Assessment of functional differences in malignant and benign breast lesions and improvement of diagnostic accuracy by using US-guided diffuse optical tomography in conjunction with conventional US. *Radiology.* 2016;280(2):387-397.
 28. Anderson RR, Parrish JA. The optics of human skin. *J Invest Dermatol.* 1981;77(1):13-9. <http://www.sciencedirect.com/science/article/pii/S0022202X15461251>.
 29. Jacques JA, Huss J, McKeenan W, Dimitroff JM, Kuppenheim HF. Spectral reflectance of human skin in the region 0.7–2.6 μ . *J Appl Physiol.* 1955;8(3):297-299. <http://jap.physiology.org/content/8/3/297.abstract>.
 30. Mendenhall MJ, Nunez AS, Martin RK. Human skin detection in the visible and near infrared. *Appl Opt.* 2015;54(35):10559-10569. doi:10.1364/AO.54.010559.
 31. Chen NG, Guo P, Yan S, Piao D, Zhu Q. Simultaneous near-infrared diffusive light and ultrasound imaging. *Appl Opt.* 2001;40(34):6367-6380.
 32. Boas D, Gaudette T, Arridge S. Simultaneous imaging and optode calibration with diffuse optical tomography. *Opt Express.* 2001;8(5):263-270.
 33. Tarvainen T, Kolehmainen V, Vauhkonen M, et al. Computational calibration method for optical tomography. *Appl Opt.* 2005;44(10):1879-1888.
 34. Stott JJ, Culver JP, Arridge SR, Boas DA. Optode positional calibration in diffuse optical tomography. *Appl Opt.* 2003;42(16):3154-3162. <http://ao.osa.org/abstract.cfm?URI=ao-42-16-3154>.
 35. Schweiger M, Nissilä I, Boas DA, Arridge SR. Image reconstruction in optical tomography in the presence of coupling errors. *Appl Opt.* 2007;46(14):2743-2756.
 36. Mozumder M, Tarvainen T, Arridge SR, Kaipio J, Kolehmainen V. Compensation of optode sensitivity and position errors in diffuse optical tomography using the approximation error approach. *Biomed Opt Express.* 2013;4(10):2015-2031.
 37. Fukuzawa R, Hoshi Y, Fukuzawa R, et al. Reduction of image artifacts induced by change in the optode coupling in time-resolved diffuse optical tomography in the optode coupling in time-resolved diffuse. *J Biomed Opt.* 2011;16(11):116022.
 38. Xu H, Pogue BW, Springett R, Dehghani H. Spectral derivative based image reconstruction provides inherent insensitivity to coupling and geometric errors. *Opt Lett.* 2005;30(21):2912-2914.
 39. Mozumder M, Tarvainen T, Seppänen A, Nissilä I, Arridge SR, Kolehmainen V. Nonlinear approach to difference imaging in diffuse optical tomography. *J Biomed Opt.* 2015;20(10):105001. <http://biomedicaloptics.spiedigitallibrary.org/article.aspx?doi:10.1117/1.JBO.20.10.105001>.
 40. Vavadi H, Zhu Q. Automated data selection method to improve robustness of diffuse optical tomography for breast cancer imaging. *Biomed Opt Express.* 2016;7(10):4007-4020. <http://www.osapublishing.org/boe/abstract.cfm?URI=boe-7-10-4007>.
 41. Fisher SRA. *Statistical Methods for Research Workers.* 13th ed. Edinburgh, Scotland; London, England: Oliver and Boyd; 1958.
 42. Kendall MG. *The Advanced Theory of Statistics.* 4th ed. London, England: Macmillan; 1979.
 43. Grubbs FE. Sample criteria for testing outlying observations. *Ann Math Stat.* 1950;21(1):27-58.
 44. Grubbs FE. Procedures for detecting outlying observations in samples. *Technometrics.* 1969;11(1):1-21. <http://www.tandfonline.com/doi/abs/10.1080/00401706.1969.10490657>.
 45. Cope M. *The Development of a Near Infrared Spectroscopy System and Its Application for Non Invasive Monitoring of Cerebral Blood and Tissue Oxygenation in the Newborn Infants.* London, England: University of London; 1991.