

Near-Infrared Parameters Extraction: A Potential Method To Detect Skin Cancer

Bao C. Q. Truong*, H. D. Tuan* *Member, IEEE* and H. T. Nguyen* *Senior Member, IEEE*

Abstract—The wavelength-dependent absorption coefficients can be used to analyse optical properties of human skin. Existing absorption models for narrow ranges in the visible and near infrared are insufficient to simultaneously incorporate the spectral contrast produced by differences in chromophores, water and lipid content of skin tissue into skin cancer detection. In the broad range up to 1600 nm, recent analysis approaches for absorption spectra do not consistently provide significant differences between healthy and cancerous skins. We propose an absorption model to fit the absorption coefficient spectra of skin samples over the range from 400 nm to 1600 nm and an advanced algorithm to find the optimal estimation. The extracted parameters of this model are analysed by a statistical t-test. The test results demonstrate the significant differences between all pairs of tumour-normal skin. Therefore, our approach has strong potential for early skin cancer detection using near infrared spectroscopy (NIRS).

I. INTRODUCTION

Near-infrared spectroscopy is inexpensive and potential for both diagnosis and therapeutic applications of many diseases, especially cancer in brain, breast, cervix, skin, prostate and pancreas [1]–[3]. Indeed, optical properties of tissue reveal its structural information such as thickness, composition, morphology, chromophore contents and layer structures [4], [5]. Various quantitative models have been developed to approximate the absorption spectra of human skin in the spectral range below 1000 nm [6]–[8]. Physiological parameters of these models such as chromophore concentrations and chemical contents constitute the investigation of skin response to near-infrared light. The wavelength-dependent absorption coefficient of human skin are extracted from reflectance and transmittance spectra [5], [9]–[11].

Absorption spectra of human skin have been studied in local narrow ranges of wavelength [8], [10], [12]–[14]. Particularly, Reference [13] the substantial contribution of melanin and haemoglobin to total absorption spectra of human skin in the range from 460 to 820 nm. Furthermore, melanin revealed the differences in absorption properties between melanoma and dysplastic nevi in the wavelength band up to 1000 nm [8]. In the wavelength range from 1000 to 2200 nm, the reflectance and transmittance spectra were measured in many samples of skin layers, i.e. stratum corneum, epidermis and dermis [10], which supports the role of water and lipid in human skin as the main absorbers in the spectral region. The differences in water and lipid content between normal and cancerous skin result in their spectral contrast which were presented in [15] in the

range from 1000 to 1500 nm. In fact, absorption spectra in a local narrow range below 1000 nm are not accountable for the impacts of water and lipid. Therefore, the absorption spectra corresponding to the wavelength range from the visible range to near-infrared beyond 1000nm can improve the contrast between healthy and cancerous skin. The existing approach for analysing these spectra is based on spectral visualization but its outcomes are still by far applicable for a clinical application [15], [16]. Indeed, Reference [15]’s results demonstrate a lower accuracy than those by biopsies while the contrast in absorption spectra investigated by [16] between tumour and normal skin is inconsistent.

The present study aims to address the challenges aforementioned by [15], [16]. We develop a proper mathematical model to estimate the absorption spectra of human skin. An advanced global optimization algorithm is applied to obtain the optimal estimation. We investigate the extracted parameters of our model using a Student’s t-test described later in order to analyse their potential for differentiating between normal skin and nonmelanoma skin cancer (NMSC). The simulation and analysis results confirm the feasibility of our approach to detect skin cancer.

The paper structure is as follows. Section II is devoted to modelling the light absorption of human skin. Section III provides our global optimization technique to extract parameters and the method of statistical analysis. The simulation results and discussion are given in Section IV. Conclusion is in Section V.

II. ABSORPTION MODELLING

According to the Beer-Lambert law, a typical model to describe the attenuation of light propagating inside the scattering media is constructed by the linear approximation of known chromophore concentrations and their wavelength-dependent extinction coefficients. The absorption caused by haemoglobin and deoxyhaemoglobin is dominant in the wavelength range from 400 nm to 900 nm [17]. Another chromophore contributing to light absorption of skin is melanin of which absorbance behavior has been well-described by a wavelength-dependent exponential function [8]. In addition, the absorption spectra in the wavelength range from 900 nm to 2000 nm are mainly dominated by water and lipid absorbance [10]. Therefore, it is necessary to simultaneously involve strong absorbers including haemoglobin, deoxyhaemoglobin, melanin, water and lipid into describing the absorption spectra of skin tissue in the wavelength range from 400 nm to 1600 nm. Specifically, we can calculate the absorption coefficient $\mu_A(\lambda_i)$ of human skin

*Centre for Health Technologies, FEIT, University of Technology, Sydney, AUSTRALIA; Email: cao.q.truong@student.uts.edu.au, tuan.hoang@uts.edu.au, hung.nguyen@uts.edu.au

at a wavelength λ_i in this wavelength domain as follows:

$$\begin{aligned} \mu_A(\lambda_i) &= \mathbf{c}_{Hb}[(1 - \boldsymbol{\alpha})e_{Hb}(\lambda_i) + \boldsymbol{\alpha}e_{HbO_2}(\lambda_i)] \\ &+ \mathbf{c}_m \exp\left[-\mathbf{k} \frac{\lambda_i - \lambda_0}{\lambda_0}\right] + \mathbf{c}_{H_2O}[\epsilon_{H_2O}(\lambda_i) + \mathbf{r}\epsilon_{lipid}(\lambda_i)]. \end{aligned} \quad (1)$$

Here $e_{HbO_2}(\lambda_i)$ and $e_{Hb}(\lambda_i)$ are the extinction coefficients of deoxy-haemoglobin and haemoglobin at λ_i (cm^{-1}/M). $\epsilon_{H_2O}(\lambda_i)$ and $\epsilon_{lipid}(\lambda_i)$ are the absorption spectra of water and lipid (cm^{-1}) at λ_i respectively. \mathbf{c}_{Hb} and $\boldsymbol{\alpha}$ are the haemoglobin concentration (M) and the haemoglobin oxygen saturation of blood respectively. \mathbf{c}_m and \mathbf{k} are respectively the melanin concentration and the exponential decay constant. Based on available studies in the literature, the variation range of \mathbf{k} has been determined roughly from 1.2 to 5.5 [8]. λ_0 is the wavelength constant at which melanin absorption is the highest [8]. As our examining spectral region is from 400 to 1600 nm, we choose $\lambda_0 = 400$. \mathbf{c}_{H_2O} is the water proportion in tissue, and \mathbf{r} is the ratio of lipid to water.

III. PARAMETERS EXTRACTION PROCEDURE

A. Extraction Algorithm

The six parameters $\mathbf{x} = (\mathbf{c}_{Hb}, \boldsymbol{\alpha}, \mathbf{c}_m, \mathbf{k}, \mathbf{c}_{H_2O}, \mathbf{r})$ are extracted by minimizing the total square error function of $\mu_A(\lambda_i)$ and its corresponding measured value $\mu_{Am}(\lambda_i)$ over all the wavelength samples in the range from 400 to 1600 nm.

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{i=1}^N ER_A(\lambda_i) = \sum_{i=1}^N |\mu_A(\lambda_i) - \mu_{Am}(\lambda_i)|^2 \\ \text{subject to} \quad & \mathbf{c}_{Hb} \geq 0, 0 \leq \boldsymbol{\alpha} \leq 1, \mathbf{c}_m \geq 0, \\ & 1.2 \leq \mathbf{k} \leq 5.5, \\ & 0 \leq \mathbf{c}_{H_2O} \leq 1, \mathbf{r} \geq 0. \end{aligned} \quad (2)$$

Here N is the total number of sampled wavelengths. Obviously, the three variables $(\boldsymbol{\alpha}, \mathbf{k}, \mathbf{r})$ make $ER_A(\lambda_i)$ nonlinear and nonconvex [18]. Therefore, we consider the following variables changes,

$$\bar{\mathbf{k}} = \mathbf{k}/\lambda_0, \bar{\mathbf{c}}_m = \mathbf{c}_m \exp(\mathbf{k}), \bar{\boldsymbol{\alpha}} = \mathbf{c}_{Hb}\boldsymbol{\alpha}, \bar{\mathbf{r}} = \mathbf{c}_{H_2O}\mathbf{r}. \quad (3)$$

Consequently, $\mu_A(\lambda_i)$ is simplified to

$$\mu_A(\lambda_i) = a_i^T \bar{\mathbf{x}} \quad (4)$$

with

$$a_i = \begin{bmatrix} e_{Hb}(\lambda_i) \\ e_{HbO_2}(\lambda_i) - e_{Hb}(\lambda_i) \\ \exp(-\lambda_i \bar{\mathbf{k}}) \\ \epsilon_{H_2O}(\lambda_i) \\ \epsilon_{lipid}(\lambda_i) \end{bmatrix}, \bar{\mathbf{x}} = \begin{bmatrix} \mathbf{c}_{Hb} \\ \bar{\boldsymbol{\alpha}} \\ \bar{\mathbf{c}}_m \\ \mathbf{c}_{H_2O} \\ \bar{\mathbf{r}} \end{bmatrix}. \quad (5)$$

Furthermore, if $\bar{\mathbf{k}}$ is held fixed, $ER_A(\lambda_i)$ becomes a convex function in $\bar{\mathbf{x}}$. This observation results in converting (2) into the following parametric optimization,

$$\min_{\bar{\mathbf{k}}} F(\bar{\mathbf{x}}) : 0.003 \leq \bar{\mathbf{k}} \leq 0.0138 \quad (6)$$

with

$$\begin{aligned} F(\bar{\mathbf{x}}) := \min_{\bar{\mathbf{x}}} \quad & \sum_{i=1}^N ER_A(\lambda_i) : \mathbf{c}_{Hb} \geq 0, \bar{\boldsymbol{\alpha}} \geq 0, \bar{\mathbf{c}}_m \geq 0, \\ & 0.003 \leq \bar{\mathbf{k}} \leq 0.0138, 0 \leq \mathbf{c}_{H_2O} \leq 1, \mathbf{c}_{H_2O} + \bar{\mathbf{r}} \leq 1. \end{aligned} \quad (7)$$

Like [19] we employ the grinding optimization to solve (6). To be more specific, we choose M values of $\bar{\mathbf{k}}$ based on its constraint, and problem (7) is solved at each $\bar{\mathbf{k}}(j) := 0.003 + j(0.0138 - 0.003)/M$. The optimal solution of (6) is determined with the $\bar{\mathbf{k}}(j)$ which contributes to the minimum objective value of problem (7). Let $\bar{\mathbf{x}} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)^T$. Problem (7) is performed as the following convex quadratic program in five scalar variables $\bar{\mathbf{x}} \in \mathbf{R}^5$,

$$\begin{aligned} \min_{\bar{\mathbf{x}} \in \mathbf{R}^5} \quad & \sum_{i=1}^N |a_i^T \bar{\mathbf{x}} - \mu_{Am}(\lambda_i)|^2 = \bar{\mathbf{x}}^T A \bar{\mathbf{x}} + b^T \bar{\mathbf{x}} + d \\ \text{subject to} \quad & \mathbf{x}_1 \geq 0, \mathbf{x}_2 \geq 0, \mathbf{x}_3 \geq 0, 0 \leq \mathbf{x}_4 \leq 1, \\ & \mathbf{x}_5 \geq 0, \mathbf{x}_4 + \mathbf{x}_5 \leq 1 \end{aligned} \quad (8)$$

where

$$A = \sum_{i=1}^N a_i^* a_i^T, b^T = \sum_{i=1}^N (-2\mu_{Am}(\lambda_i) a_i^T), d = \sum_{i=1}^N (\mu_{Am}(\lambda_i))^2.$$

Problem (8) is solved by SeduMi [20] corresponding to each $\bar{\mathbf{k}}(j)$. Conventionally, \mathbf{x} is recovered from $\bar{\mathbf{x}}$ thanks to (3).

B. Data Accumulation

Haemoglobin and deoxyhaemoglobin extinction coefficient spectra were published by Oregon Medical Laser Centre (OMLC) [21]. The absorption spectra of water and lipid were extracted from [22]. The data used to examine our model and algorithm have been presented in [16]¹. Particularly, there are six types of skin samples examined in [16] including epidermis, dermis, subcutaneous fat, infiltrative basal cell carcinomas (BCC), nodular BCC, and squamous cell carcinomas (SCC). These specimens were excised from the face, neck, and back of patients. Absorption spectra of these human skins were extracted from reflectance and transmittance spectra in the wavelength range from 400 to 1600 nm with their corresponding standard errors.

C. Statistical Analysis

We employ the two-sample t-test for pairs of skin groups, especially normal and cancerous tissues, to assess the level of difference in their extracted parameters from (1) [23]. In the test, the standard errors of the measured absorption spectrum of each skin sample are utilised to determine the variation range of each model parameter corresponding to that sample. The statistical hypothesis value *p-value* is to evaluate the significance of difference between extracted parameters of skin samples in statistics perspective. The significant difference between two samples is confirmed by their *p-value* ≤ 0.05 . Furthermore, the lower value the *p-value* of a t-test reaches to, the higher the possibility of distinguishing one sample from another is.

¹We would like to thank Dr. Yaroslavsky, who is a co-author of [16], for providing the data.

IV. RESULTS AND DISCUSSION

The absorption spectra are fitted by our proposed model within the range from 400 to 1600 nm. We used 270 wavelength samples ($N = 270$) in the spectral region for each case. The number of grinding points M applied to the extraction algorithm are chosen as 5000. Figs.1-3 reveal the potential of our absorption model to describe the absorbance of skin tissue since the simulated spectra by the model (1) fit the real absorption spectra very well. We also analyse the fits in terms of statistics. In particular, the goodness of fitting (GOF) is assessed through the $R - square$ which indicates the fitting quality of numerical models (see e.g. [24]). A perfect fitting is represented by $R - square = 1$. The $R - square$ of each fitting was provided in Table I. According to the values of $R - square$, our proposed optimization algorithm based on grinding method is able to locate the global minimum of the error function (2). This outcome facilitates the investigation of changes in haemoglobin, melanin and water contents of skin tissue over a wide wavelength range. The particular values of the extracted parameters corresponding to each skin sample are listed in Table I. The quantitative analysis based on the results of our extraction demonstrated an expected agreement with the conclusions of [16]. Indeed, water absorbance is the main factor contributing to the absorption spectra of nonmelanoma skin in the spectral range above 900 nm, which is apparently recognized through the hydration proportion c_{H_2O} of the samples. Furthermore, the water proportion in subcutaneous fat is considerably lower than those in nonmelanoma skins. On the other hand, the amount of blood in nodular BCC is smaller than one in infiltrative BCC and SCC according to the haemoglobin concentration c_{Hb} of these specimens.

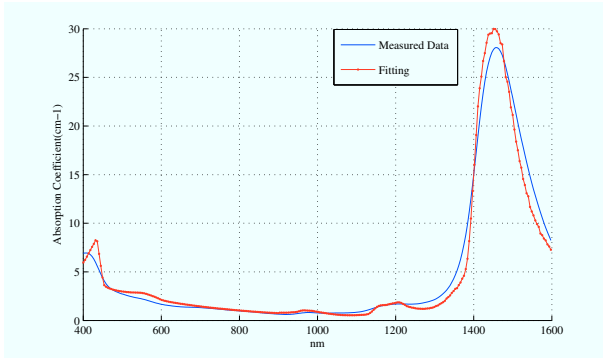


Fig. 1. The absorption coefficients of infiltrative BCC: the measured data of [16] and the fitting curve.

The significance of differences between normal skin and NMSC were examined by p -value of the applied t-test in Table II. The p -values which are smaller than the critical value 0.05 are highlighted in bold type. It is obvious that each pair of healthy skin and tumor is distinguishable by using their respective extracted parameters. More importantly, our approach even provided higher possibility to differentiate between cancerous and normal skins than [16]. In detail, while the spectra-based analysis of [16] failed to reveal any wavelength region consistently providing significant differences for all pairs of cancer-normal tissue, our parameter-based analysis

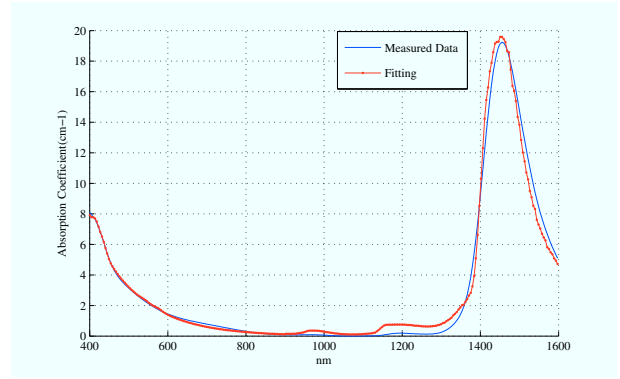


Fig. 2. The absorption coefficients of nodular BCC: the measured data of [16] and the fitting curve.

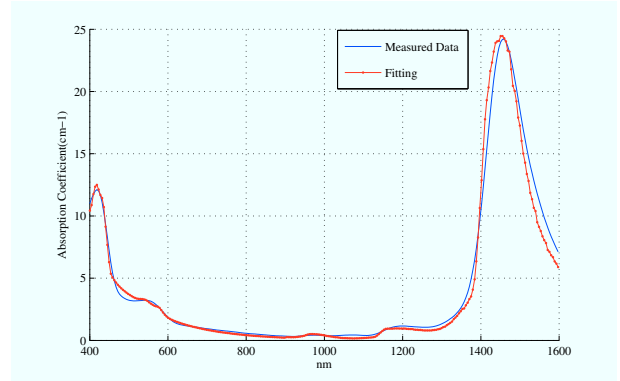


Fig. 3. The absorption coefficients of SCC: the measured data of [16] and the fitting curve.

is successful to provide the stable results of the t-test. Indeed, all p -values of c_m and c_{H_2O} are smaller than 0.05. Therefore, the extracted c_m and c_{H_2O} presented a consistent contrast between healthy skin and tumour. These outcomes support using water proportion c_{H_2O} and melanin concentration c_m in tissue to detect skin cancer.

V. CONCLUSION

Our first contribution is proposing a mathematical model successfully describing the absorption property of human skin in NIR range of wavelength. Secondly, the extracted parameters by our global optimization algorithm are examined by the two-sample t-test, and their significant differences appear in all pairs of tumour-normal skin. These results show potential for differentiating between non-melanoma and normal skin. Furthermore, the consistent differences in the water proportion and melanin concentration between tumour and healthy skin are worth considering as an crucial source of contrast to develop an advanced classification in future studies. In further research, the scattering property of human skin shall also be investigated with our parameter-based approach to improve the possibility of detecting skin cancer with NIR.

REFERENCES

- [1] V. Kondepoti, H. Heise, and J. Backhaus, "Recent applications of near-infrared spectroscopy in cancer diagnosis and therapy," *Anal. Bioanal. Chem.*, vol. 390, pp. 125–139, 2008.

Table I
EXTRACTED PARAMETERS OF THE ABSORPTION MODEL (1)

Sample	c_{Hb}	α	c_m	k	c_{H_2O}	r	$R - square$
Infiltrative BCC	9.074908	0.648784877	3.473577002	1.20056	0.91113571	0.097531331	0.971053961
Nodular BCC	2.744060	0.616242678	7.239505634	3.32128	0.599445617	2.83273E-09	0.98731347
SCC	14.633063	0.431173452	6.861462813	2.81112	0.748247297	1.7744E-08	0.973738938
Epidermis	2.153532	0.51736626	12.35606729	2.39616	0.957576863	4.16735E-09	0.987869874
Dermis	7.138897	0.636322217	6.468469829	2.2248	0.712403753	0.403696851	0.978049291
Subcutaneous Fat	13.949501	0.622986038	12.24294198	3.3252	0.350454799	1.853435027	0.940927763

Table II
 p -value CORRESPONDING TO EXTRACTED PARAMETERS OF THE ABSORPTION MODEL (1)

$P - value$	c_{Hb}	α	c_m	k	c_{H_2O}	r
Epidermis-InfBCC	0.000292809	0.03363132	4.27E-05	0.00038	0.036995	0.02016
Epidermis-Nod BCC	0.053133051	0.037978	0.000484	8.93E-06	0.0001	0.0837
Epidermis-SCC	3.3E-06	0.082	7.46E-05	0.003	0.00027	0.0433
Dermis-InfBCC	0.00049	0.00098	2.976E-05	0.00158	0.0011	6.51E-05
Dermis-Nod BCC	6.05E-05	0.257	0.015	7.85E-05	2.9E-05	0.00011
Dermis-SCC	0.0002	2.55E-06	0.00705	9.55E-05	0.0029	0.00011
Fat-InfBCC	0.0043	0.018	7.65E-08	3.8E-05	0.00015	1.96E-05
Fat-Nod BCC	0.0008	0.652675619	3.33E-05	0.908	9.47E-07	0.00025
Fat-SCC	0.1684	2.65E-05	4.2E-07	0.00051	2.06E-07	0.00025

- [2] M. Ferrari, L. Mottola, and V. Quaresima, "Principles, techniques, and limitations of near infrared spectroscopy," *Can. J. Appl. Physiol.*, vol. 29, no. 4, pp. 463–487, 2004.
- [3] V. R. Kondepoti, M. Keese, R. Mueller, B. C. Manegold, and J. Backhaus, "Application of near-infrared spectroscopy for the diagnosis of colorectal cancer in resected human tissue specimens," *Vibrational Spectroscopy*, vol. 44, no. 2, pp. 236 – 242, 2007.
- [4] R. R. Anderson and A. J. Parrish, "The optics of human skin," *J. Invest. Dermatol.*, vol. 1, no. 77, pp. 13–19, 1981.
- [5] W. Cheong, S. Prahl, and A. Welch, "A review of the optical properties of biological tissues," *IEEE J. Quantum Electronics*, vol. 26, no. 12, pp. 2166 –2185, 1990.
- [6] S.-H. Tseng, P. Bargo, A. Durkin, and N. Kollias, "Chromophore concentrations, absorption and scattering properties of human skin in vivo," *Opt. Express*, vol. 17, no. 17, pp. 14599–14617, 2009.
- [7] G. Zonios and A. Dimou, "Modeling diffuse reflectance from semi-infinite turbid media: Application to the study of skin optical properties," *Opt. Express*, vol. 14, pp. 8661–8674, Sep 2006.
- [8] G. Zonios, A. Dimou, I. Bassukas, D. Galaris, A. Tsolakidis, and E. Kaxiras, "Melanin absorption spectroscopy: new method for non-invasive skin investigation and melanoma detection," *J. Biomed. Opt.*, vol. 13, no. 1, 2008.
- [9] S. Prahl, *Light transport in tissue*. Phd dissertation, University of Texas at Austin, 1988.
- [10] T. L. Troy and S. N. Thennadil, "Optical properties of human skin in the near infrared wavelength range of 1000 to 2200 nm," *J. Biomed. Opt.*, vol. 6, no. 2, pp. 167–176, 2001.
- [11] L. Wang, S. L. Jacques, and L. Zheng, "MCML-Monte Carlo modeling of light transport in multi-layered tissues," *Comput Methods Programs Biomed.*, vol. 47, no. 2, pp. 131 – 146, 1995.
- [12] C. R. Simpson, M. Kohl, M. Essenpreis, and M. Cope, "Near-infrared optical properties of ex vivo human skin and subcutaneous tissues measured using the monte carlo inversion technique," *Phys. Med. Bio.*, vol. 43, no. 9, p. 2465, 1998.
- [13] G. Zonios, J. Bykowski, and N. Kollias, "Skin melanin, hemoglobin, and light scattering properties can be quantitatively assessed in vivo using diffuse reflectance spectroscopy," *J. Invest. Dermatol.*, vol. 117, no. 6, 2001.
- [14] S.-H. Tseng, A. Grant, and A. J. Durkin, "In vivo determination of skin near-infrared optical properties using diffuse optical spectroscopy," *J. Biomed. Opt.*, vol. 13, no. 1, p. 014016, 2008.
- [15] L. M. McIntosh and R. Summers, "Towards non-invasive screening of skin lesions by near-infrared spectroscopy," *J. Invest. Dermatol.*, vol. 116, pp. 175–181, 2001.
- [16] E. Salomatina, B. Jiang, J. Novak, and A. N. Yaroslavsky, "Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range.," *J. Biomed. Opt.*, vol. 11, p. 064026, 2006.
- [17] A. N. Bashkatov, E. A. Genina, V. I. Kochubey, and V. V. Tuchin, "Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm," *J. Phys. D: Appl. Phys.*, vol. 38, no. 9, 2005.
- [18] H. Tuy, *Convex Analysis and Global Optimization*. Kluwer Academic, 1998.
- [19] B. C. Q. Truong, H. D. Tuan, H. H. Kha, and H. T. Nguyen, "System identification for Terahertz wave's propagation and reflection in human skin," in *2012 Fourth International Conference on Communications and Electronics*, pp. 364 –368, 2012.
- [20] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optim. Method. Softw.*, vol. 11, no. 1, pp. 625–653, 1999.
- [21] S. Prahl, "Optical absorption of hemoglobin," *Oregon Medical Laser Center*, <http://omlc.ogi.edu/spectra/hemoglobin/index.html>, 1999.
- [22] R. Nachabe, B. H. W. Hendriks, A. E. Desjardins, M. van der Voort, and M. van der Mark, "Estimation of lipid and water concentrations in scattering media with diffuse optical spectroscopy from 900 to 1600 nm," *J. Biomed. Opt.*, vol. 15, no. 3, 2010.
- [23] MATLAB, *Statistics Toolbox*. The MathWorks Inc.: Natick, MA, U.S.A., 2012.
- [24] MATLAB, *Curve Fitting Toolbox*. The MathWorks Inc.: Natick, MA, U.S.A., 2012.