Skin surface detection in 3D optoacoustic mesoscopy based on dynamic programming

Suhanyaa Nitkunanantharajah, Guillaume Zahnd, Malini Olivo, Nassir Navab, Pouyan Mohajerani and Vasilis Ntziachristos*

Abstract — **Optoacoustic (photoacoustic) mesoscopy offers unique capabilities in skin imaging and resolves skin features associated with detection, diagnosis and management of disease. A critical first step in the quantitative analysis of clinical optoacoustic images is to identify the skin surface in a rapid, reliable and automated manner. Nevertheless, most common edge- and surface-detection algorithms cannot reliably detect the skin surface on 3D raster-scan optoacoustic mesoscopy (RSOM) images, due to discontinuities and diffuse interfaces in the image. We present herein a novel dynamic programming approach that extracts the skin boundary as a 2D surface in one single step, as opposed to consecutive extraction of several independent 1D contours. A domain-specific energy function is introduced, taking into account the properties of volumetric optoacoustic mesoscopy images. The accuracy of the proposed method is validated on scans of the volar forearm of 19 volunteers with different skin complexions, for which the skin surface has been traced manually to provide a reference. Additionally, the robustness and the limitations of the method are demonstrated on data where the skin boundaries are low-contrast or ill-defined. The automatic skin surface detection method can improve the speed and accuracy in the analysis of quantitative features seen on RSOM images and accelerate the clinical translation of the technique. Our method can likely be extended to identify other types of surfaces in RSOM and other imaging modalities.**

Index Terms—**2D front propagation, optoacoustic imaging, skin extraction, surface segmentation**

I. INTRODUCTION

DERMAL morphology has high diagnostic value [1, 2]. Changes in microvascular structure and in the appearance and thickness of skin layers have been shown to be useful for assessing different pathologies, including skin and systemic diseases [3–9]. Optical imaging is widely used in dermatology for high-resolution visualization of skin features. However, optical imaging and microscopy come with penetration depths of tens to a few hundreds of microns and are therefore not well suited for assessing dermal biomarkers [10]. Optical coherence tomography (OCT), is an alternative method that visualizes photon scattering contrast in tissues and produces high-resolution images at penetration depths of ~1 mm. However, OCT lacks sensitivity to tissue chromophores [11–13]. Ultrasonography has greater penetration depth, but suffers from reduced contrast [14]. Optoacoustic (photoacoustic) imaging bridges the gap between optical and acoustic imaging modalities. It images tissue at high resolution and specificity from optical contrast, while penetrating deep into the sample [14]. Ultra-wideband raster-scan optoacoustic mesoscopy (RSOM) [15, 16] has recently demonstrated high-resolution skin imaging by revealing different skin layers and the structure of the microvasculature [17, 18]. The method has been used for in-depth visual examination of psoriasis [5] and analysis of vascularization of superficial tumors [19]. Previous work [19] showed that detecting an extended ultrasound bandwidth improves imaging of fine anatomical details such as small vessels and capillary loops.

The rich, three-dimensional information contained in RSOM images can be employed to compute skin features and biomarkers such as the total blood volume, the thickness of the epidermis and the density and the diameter of the capillary loops, which can, for instance, help to assess disease progression and identify skin inflammation [5]. However, so far, assessment of these imaging biomarkers has been

S. Nitkunanantharajah, P. Mohajerani and V. Ntziachristos are with the Chair of Biological Imaging and TranslaTUM, Technische Universität München, Munich, Germany and with the Institute of Biological and Medical Imaging, Helmholtz Zentrum München, Neuherberg, Germany. (E-Mail: v.ntziachristos@tum.de)

G. Zahnd is with the Chair of Computer Aided Medical Procedures, Technische Universität München, Munich, Germany.

M. Olivo is with the Lab of Bio-Optical Imaging, Singapore Bioimaging Consortium, Agency for Science Technology and Research (A*STAR), Singapore.

N. Navab is with the Chair of Computer Aided Medical Procedures, Technische Universität München, Munich, Germany and Johns Hopkins University, Baltimore, MD, USA.

* Correspondence to Vasilis Ntziachristos (E-Mail: v.ntziachristos@tum.de).

performed manually [5]. To clearly separate skin layers from each other and assess the thickness and structure of microvasculature in each layer, knowledge of the exact skin location in the data is required, or alternatively volumes with flattened skin surfaces must be generated. Accurate determination of the skin surface is a key remaining limitation that hinders the automatic detection of skin features and their quantitative analysis. Currently, the skin boundary is approximated by manual tracings performed in a few slices of the scanned volume. The manually assigned points on the skin are then used to interpolate a complete surface extending over the entire scanned region. This type of manual procedure is tedious, time-consuming, and user-dependent. Furthermore, the accuracy of such annotations depends on the smoothness of the skin surface and the number of slices selected. RSOM images frequently contain an uneven or tilted surface, which compromises the segmentation accuracy. Therefore, automation and standardization of skin surface detection would increase the accuracy and speed of RSOM image processing and facilitate its clinical translation.

RSOM is most often used to image hemoglobin, so the images depict primarily the network of subcutaneous vessel structures rather than the continuous surface of the stratum corneum. Therefore, RSOM does not provide a closed, continuous, well-defined interface. The 2D maximum intensity projections (MIPs) of RSOM volumes usually reveal the skin boundary, but within the volume itself, the boundary can suffer from local discontinuities and low-contrast interfaces.

An initial approach for skin boundary detection has been proposed for optoacoustic microscopy [20], where the skin location is roughly estimated by analyzing the signal amplitudes of each A-line, followed by outlier removal using regression smoothing. The resulting estimated surface is subjected to low-pass filtering, generating a smooth 2D skin profile. However, multiple smoothing steps in the extraction process impair the accuracy of the detection and provide only an estimate of the skin location. Furthermore, the algorithm requires two local regression smoothing steps, which are computationally intensive.

A more efficient segmentation method without additional smoothing steps was previously proposed [21] to segment the fibrous cap in OCT volumes. The segmentation framework is based on the dynamic programming (DP) strategy and uses front propagation to successfully determine boundaries, even in signal-poor regions with ill-defined and blurred interfaces. An expansion of the DP-based segmentation method to 3D volumes has been demonstrated with sequential magnetic resonance imaging (MRI) [22]. First, the volume is sliced in one direction and an initial round of 1D front propagation is performed on each slice. The propagated front is not used to directly extract the contours; instead, the cumulative cost values are used to generate cost values in preparation for a second propagation. Subsequently, a second round of 1D front propagations is performed on 2D slices of the cumulative costs across the orthogonal direction. Although this method manages to detect a surface even over discontinuities of the contour, it is strongly direction-dependent, as it usually produces different results depending on the first and second directions in which the volumes are sliced.

For accurate skin boundary recognition in RSOM images, a 3D segmentation method is needed that can extract smooth surfaces in a direction-independent manner despite discontinuities and interfaces with low contrast. This method should work reliably and accurately without computationally intensive smoothing steps in order to detect the skin surface in RSOM volumes.

We hypothesized that if we extended DP-based segmentation methods to perform front propagation in two dimensions simultaneously and directly extracted a 2D surface, as opposed to two independent 1D front propagations, then we could extract the skin boundary in 3D optoacoustic mesoscopic volumes efficiently even over discontinuities in the skin contour without any smoothing steps.

In this work, we introduce a DP-based segmentation method using 2D front propagation in combination with backtracking for surface extraction. The main contribution of this work is the ability to extract the entire 2D surface in one step. In order to properly detect the skin surface displayed in RSOM volumes, an RSOM-specific energy function was developed that takes into consideration the occurrence of RSOM-specific surface-like artifacts as well as the discontinuous, ill-defined interfaces. The accuracy of the method was validated on RSOM scans of healthy volunteers with different skin complexions based on Fitzpatrick skin phototypes [23]; the automatic segmentation by the algorithm was compared with manual tracings. Finally, the robustness and the limitations of the method were evaluated by testing the algorithm on data with low contrast and ill-defined interfaces.

II. METHODS

In the following sections, we first give a short overview of the imaging modality and the challenges that need to be faced when extracting the skin surface. Afterwards, the general surface detection algorithm is outlined, before each of its components is explained in detail. Subsequently, the RSOM-specific volume energy computation is described. Finally, the method used to evaluate the performance of the surface detection is outlined.

Throughout this work, volumes and 3D tensors are denoted as V (bold and italic). $U*V$ represents the 3D convolution of two tensors, while $U \cdot V$ denotes an elementwise multiplication of two tensors of the same size.

A. Raster Scan Optoacoustic Mesoscopy (RSOM)

RSOM [15, 16], acquired with the ultra-broadband imaging systems described in [5], provides images with high resolution and contrast similar to optical imaging systems while being able to penetrate the tissue as deep as a few millimeters. Examples of RSOM volumes showing the vasculature in the wrist and the arm of a volunteer can be seen in Fig. 1 (a) and (b). RSOM depicts very fine anatomical structures such as microvasculature in superficial tissue and is therefore particularly suitable for dermatological applications. Those

applications mostly either require a flattened skin surface in the image or knowledge of the exact skin location. A simple skin segmentation on the 2D MIP, either automatically or manually, is fast but also prone to inaccuracies. In particular, tilts along the projected axis can cause substantial segmentation errors. At the same time, manually extracting the skin in a 3D volume slice-by-slice is generally tedious, time-consuming, and subject to inter- and intra-analyst variability. Hence, an automatic extraction of the skin surface in a 3D volume is necessary. Such an approach will enable automated analysis of large amounts of RSOM volumes and facilitate the usage of RSOM for several dermatological applications.

However, RSOM volumes do not necessarily image the skin surface itself but rather visualize the vasculature in the skin as well as the melanin layer. Consequently, they do not contain a smooth skin surface that can easily be extracted. Thus, the main challenge is to estimate the location of the skin boundary based on the visualized vasculature and extract a smooth surface on top of these irregular structures and inconsistent contours (see ROI 1 in Fig. 1 (c)). Additionally, the structures of the uppermost layer, especially the capillary loops (visualized as small green dots in Fig. 1 (b)), can have a very low contrast (see ROI 2 in Fig. 1 (d)), making it difficult to distinguish them from the noise above the skin. Meanwhile, lower structures such as prominent vessels are highly absorbent and consequently have a very high contrast. Therefore, they are more likely to be falsely detected as skin surface when using purely contrast-based approaches. In the following, we present a method that automatically estimates the location of the skin surface in the RSOM volumes, making use of RSOM-specific properties.
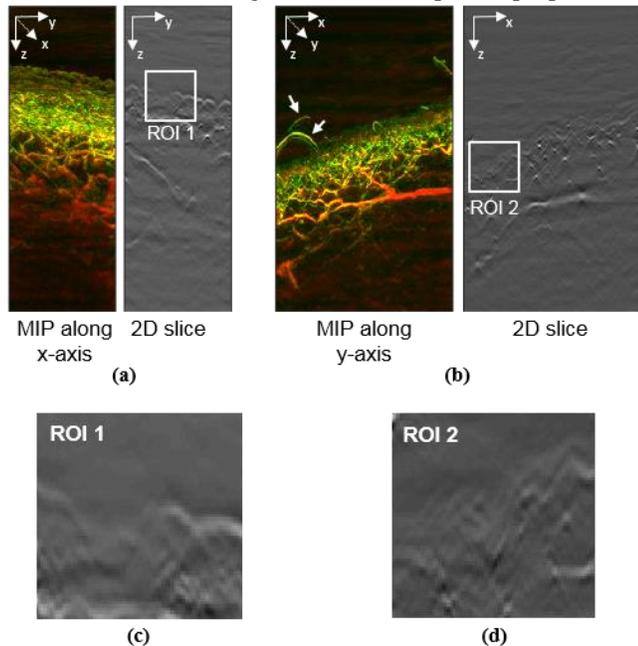

MIP along   2D slice   MIP along   2D slice
x-axis   y-axis
          (a)   (b)

          (c)   (d)

Fig. 1. Challenges of skin segmentation in optoacoustic mesoscopy. Images are shown as maximum intensity projections (MIPs) or 2D slices. **(a, b)** The skin of a healthy volunteer was imaged (a) on the wrist or (b) in mid-arm. Arrows indicate hair. Regions of interest (ROIs) are shown at higher magnification in panels (c)-(d). **(c)** Zoomed-in view of ROI 1, showing inconsistent skin contours. **(d)** Zoomed-in view of ROI 2, showing low contrast in the uppermost skin

layer.

B. Surface Detection Algorithm

The proposed method aims to extract a surface S in a 3D volume V (of size $m \times n \times o$), that passes through every (x, y)-column of V at exactly one z-position. Regarding V as a graph with every voxel representing a node and with the edges connecting a voxel to its neighborhood, the surface detection can be seen as a single-source shortest-path problem within a 3D graph [24]. In order to solve this task efficiently, the problem is split into smaller subproblems according to the dynamic programming strategy [25, 26]. Subsequently, the shortest-path problem is solved with a matrix-based approach [24] in a sequential manner. This method extends the 2D contour segmentation method using front propagation [21] by incorporating an additional spatial dimension. In this algorithm, front propagation describes an interface or "front", which moves or expands through the volume in an attempt to solve the combinatorial analysis problem by computing a shortest path for each voxel based on the known immediate neighborhood. For each voxel across the propagation directions (x,y), the method estimates the optimal depth (z) of the surface that best describe the anatomical interface. This elegant solution enables extraction of the unique optimal solution among all others, with full reproducibility and without the need to explicitly check each potential solution. The boundary surface is extracted by taking both image features and shape constraints (such as smoothness) into consideration.

The surface detection is performed by finding the surface in the volume V with the minimum cumulative cost C. The cost C is built iteratively by traversing the volume from an initial seed node using 2D front propagation and accumulating the energies of all nodes and the cost of all edges passed. The energy of a node is inversely proportional to the probability of it belonging to the surface. The cost of the connecting edges is computed based on the energy of both adjacent nodes and enforces the predefined shape constraints such as surface smoothness. In each propagation step, performed by moving either along x or y, multiple choices along z are evaluated. Once C is built, the optimal surface can be extracted by performing backtracking.

C. Cost Accumulation using 2D Front Propagation

The surface is extracted by finding the surface with the minimum cumulative cost C. Based on the DP strategy, the accumulation process is performed using 2D front propagation. The proposed method expands on a previously presented approach for 1D contour segmentation using front propagation [21, 22], but in our method the front is propagated simultaneously in an additional dimension.

For the 2D front propagation, the volume is traversed from left-to-right and from back-to-front, by increasing either x or y in each step. V is depicted as a graph and each voxel corresponds to a node; each node is characterized by a certain energy given by E, and adjacent nodes are connected by edges.
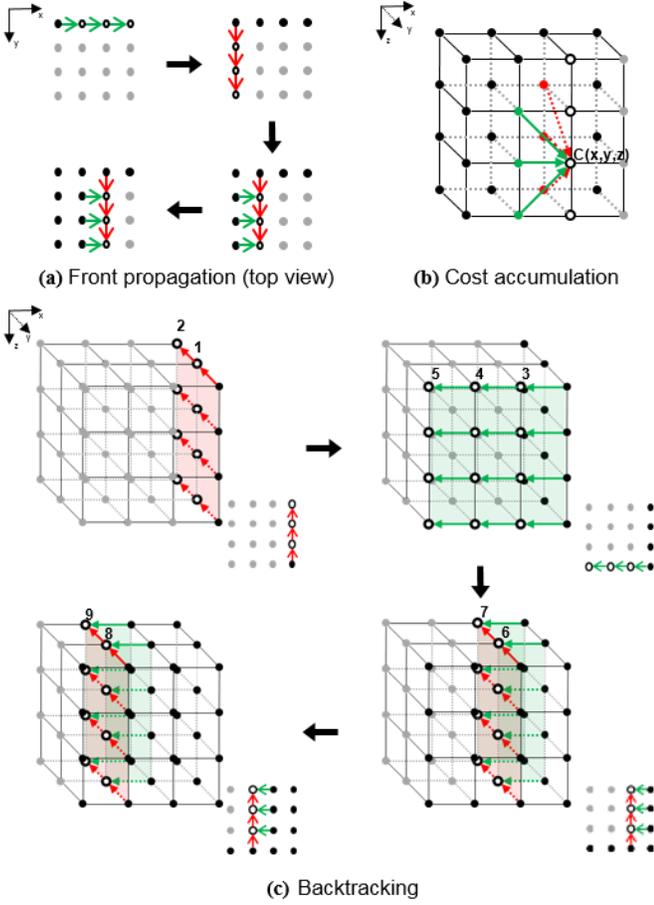
(a) Front propagation (top view)

(b) Cost accumulation

(c) Backtracking

Fig. 2 Steps of the proposed algorithm. (a) The volume traversal pattern in top view. **(b)** The cost accumulation process in the front propagation step (with d = 1). **(c)** Volume traversal during the backtracking process. Solid black points indicate nodes already computed.

The cumulative cost C for each node on the (x_i, y_i)-column is computed based on its parent nodes in the (x_{i-1}, y_i)-column and the (x_i, y_{i-1})-column (see (1)). As both parent nodes contribute equally to the cumulative cost, their contributions are averaged. For the initial nodes at depth z in the (x_0, y_0)-column, the cumulative cost is set to $E(x_0, y_0, z)$. For the border nodes that have only one parent node (either $x = x_0$ and $y \neq y_0$ or $y = y_0$ and $x \neq x_0$), the cumulative cost is fully computed based on this parent node. Fig. 2 (a) shows the volume traversal process schematically, while Fig. 2 (b) shows the nodes involved for the cost computation at a single position.

While the energy takes the features of the volume into account, the accumulation function C additionally constrains the shape of the extracted surface by penalizing steep transitions along z. The cost C for a node is computed as the sum of the cost of its parent nodes and the cost of the connecting edges between the parent nodes and the current node (as specified in

(1)). We define the energy of an edge as the average of the energies of the adjacent nodes $\left(\frac{1}{2}(E(node\ 1) + E(node\ 2)\right)$. The cost of passing an edge is determined as the energy of the edge weighted by the vertical offset. In this way, the weighting factors $(1 + \delta \cdot |i|)$ and $(1 + \delta \cdot |j|)$ control the smoothness of the extracted surface. The parameters i and j represent the vertical offsets of the surface. The neighborhood parameter d defines the adjacency of a node and, thus, limits the vertical offset. The positive parameter δ controls the smoothness of the overall surface.

When processing a given node during the front propagation process, the two parent nodes (i.e. that minimize C) along the x and y directions are stored.

D. Extraction of the Optimal Surface Using Backtracking

Backtracking starts at the ending point of the front propagation, the (x_m, y_n)-column, and traverses the volume along decreasing values of x and y. The node with the minimum cost in the (x_m, y_n)-column is the starting point of the surface. Starting from that node, the surface is built iteratively: the volume is traversed using the stored parent information, as shown in Fig. 2 (c). For each position (x_i, y_j) in the (x, y)-plane, the surface node is determined by a weighted averaging by C of the memorized parent node of the surface nodes in (x_{i+1}, y_j) and (x_i, y_{j+1}).

E. Volume Energy

The choice of the energy computation term is essential for the success of the method and should be adapted according to the task in hand.

Based on the properties of RSOM, we have designed an energy function consisting of three components: a gradient energy E_G, a surface energy E_S and a noise energy E_N. While the gradient energy gives the probability of a node of belonging to an anatomical interface, the surface energy computes its probability of belonging to the uppermost surface contained in the volume. Additionally, the noise energy helps to distinguish surface-like noise structures from the actual skin boundary. They are explained in detail in the following sections. The overall energy tensor E is then computed as the weighted sum of the individual energies (see (2)).

$$E = \alpha E_G + \beta E_S + \gamma E_N \qquad (2)$$

Each of the three energies E_G, E_S, and E_N are linearly normalized to a range of [0,1] and weighted according to the properties of the dataset. Each node in E then represents the energy of the corresponding node in V. Fig. 3 shows (a) a MIP and (b) a 2D slice of an RSOM volume V, as well as (c) the corresponding slice in E.

$$C(x,y,z) = \begin{cases} E(x,y,z) & for\ x = y = 0 \\ \min_{i \in -d...0...d}\left\{C(x-1,y,z+i) + \frac{1}{2}(E(x-1,y,z+i) + E(x,y,z))(1 + \delta \cdot |i|)\right\} & for\ y = 0 \\ \min_{j \in -d...0...d}\left\{C(x,y-1,z+j) + \frac{1}{2}(E(x,y-1,z+j) + E(x,y,z))(1 + \delta \cdot |j|)\right\} & for\ x = 0 \\ \min_{i,j \in -d...0...d}\left\{\frac{1}{2}C(x-1,y,z+i) + \frac{1}{4}(E(x-1,y,z+i) + E(x,y,z))(1 + \delta \cdot |i|) \\ + \frac{1}{2}C(x,y-1,z+j) + \frac{1}{4}(E(x,y-1,z+j) + E(x,y,z))(1 + \delta \cdot |j|)\right\} & else \end{cases} \qquad (1)$$

1) Gradient Energy E_G

Interfaces between anatomical structures manifest as regions with strong intensity changes in RSOM volumes. Those changes can be detected by filtering the volume (see (3)) with a Gaussian-weighted sign kernel P of standard deviation σ as defined in (4). σ determines how much a voxel's neighborhood influences the computation of its energy. Large values of σ help to detect boundaries better even over larger inconsistencies of the skin contour within an image plane, as multiple image slices can contribute to the gradient computation. At the same time, excessively large values of σ can cause smaller changes in the skin contour to be neglected.

$$E_G = P * V \tag{3}$$

$$P_{i,j,k} = -sgn(k)\frac{1}{(2\pi\sigma^2)^{3/2}}e^{-\frac{i^2+j^2+k^2}{2\sigma^2}} \quad \text{with} \tag{4}$$

size of $P = N \times N \times N$ and $i, j, k \in \left\{-\left\lfloor\frac{N}{2}\right\rfloor, \dots, \left\lfloor\frac{N}{2}\right\rfloor\right\}$

Fig. 3 (d) shows a 2D slice of the gradient energy E_G. The higher the changes in intensities are around a voxel, the lower is the energy in E_G, meaning greater probability that it lies on an anatomical interface.

2) Surface Energy E_S

The surface energy E_S takes the depth of a node into account and is added to the overall energy term in order to compute the probability that a node lies on the uppermost surface.

$$E_S = S \cdot (1 + S * D) \tag{5}$$

$$S(x, y, z) = \sum_{i=1}^{z} V(x, y, i) \tag{6}$$

$$D_k = -sgn(k) \quad \text{with} \tag{7}$$

size of D
$= 1 \times 1 \times M$ and $k \in \left\{-\left\lfloor\frac{M}{2}\right\rfloor, \dots, 0, \dots, \left\lfloor\frac{M}{2}\right\rfloor\right\}$

The more structures can be found above a certain voxel, the smaller is its probability of being on the uppermost surface contained in the volume. Therefore, in S the energy of each voxel v_i is computed by summing up all positive values above that voxel position (see (6)). To lower the impact of noise in this computation, edge detection is performed on S (using the kernel specified in (7)) and is added to the energy computation term of E_S as described in (5). An example for a 2D slice of the surface energy can be seen in Fig. 3 (e) (with Fig. 3 (b) showing the corresponding slice in V)

3) Noise Energy E_N

Additionally, a noise term E_N is added to the energy function representing the probability of a voxel of belonging to the RSOM-specific surface-like artifacts. Mostly occurring in regions of low signal, those artifacts emerge as flat, continuous, horizontal surfaces with homogeneous signal. The signal around the location of the skin, however, is not homogeneous. It is higher for regions encompassing vessels and lower for all other regions of the surface.

$$E_N = (\sigma_y(V) + \sigma_x(V)) * D \tag{8}$$

E_N is computed based on the sum of the standard deviation

in both x- and y-directions for each voxel (see (8)). Artifacts as well as purely noisy regions show intensity variations that are much lower than in regions containing vasculature. Therefore, the surface in between those regions can be computed by filtering the sum of standard deviations with the sign filter D described before (see (7)). Consequently, the skin surface has a low energy in E_N, whereas the surface-like artifacts have a higher energy and, therefore, are less likely to be detected as skin boundary. Fig. 3 (f) shows a slice of the E_N for the corresponding RSOM volume V as depicted in Fig. 3 (b).
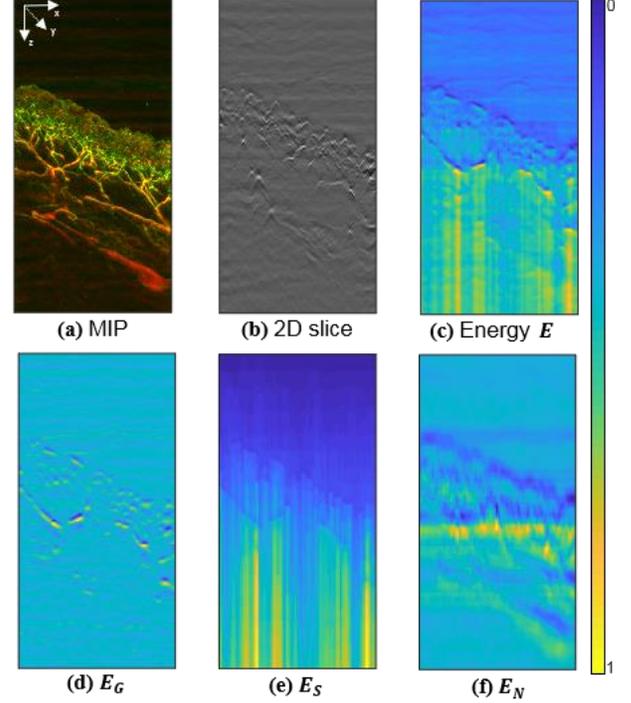

Fig. 3 Visualization of the energy computation term. (a) Maximum intensity projection (MIP) of a 3D optoacoustic mesoscopy volume V. **(b)** 2D slice of volume along the y-axis. **(c)** Slice in energy volume E. **(d)** Gradient energy E_G. **(e)** Surface penalty E_S. **(f)** Noise E_N.

F. Performance Evaluation

We evaluate our method using *in vivo* RSOM measurements, even though no real ground truth is available for such measurements. Simulated phantoms possess a ground truth but only depict a highly simplified scenario, and so cannot reflect the overall complexity of the underlying task. Generating realistic simulations would require extensive modeling of the imaging system and is out of the scope of the present work. Physical phantoms give a more realistic view of the problem but also lack a ground truth, as pressure applied to the surface of the phantom during scanning deforms the original skin boundary. Therefore, to explore the potential of our method in clinical settings, we chose *in vivo* measurements.

RSOM scans of 19 healthy volunteers of different skin complexion (with Fitzpatrick skin phototypes ranging from II to V) were used to evaluate the performance of the surface detection method. All data were acquired with the volunteers'

consent and with approval by the Domain Specific Review Board (DSRB) of the National Health Group, Singapore (2017/00932). Each volunteer was scanned at three predefined positions (wrist, midarm and elbow), resulting in 57 RSOM volumes altogether. Two data sets had to be excluded from the study after visual inspection showed the skin surface to be corrupted and only partially visible. The final test data sets contained 55 volumes.

For 11 volumes in the test data set, even manual segmentation was difficult, mainly due to large discontinuities and low contrast in the skin layer, making it difficult to distinguish the vascular structures in the upper skin layers from noise above. Nevertheless, these data were retained, classified as "difficult cases", in order to demonstrate the performance of the algorithm compared to manual segmentation. The remaining 44 volumes were classified as "standard cases".

Hair in the imaged region could potentially complicate or even obstruct the detection process. Another factor that may have an impact on the detection accuracy of the algorithm are highly tilted skin surfaces. In order to assess the robustness and the limitations of the proposed method, data with hair and with tilted surfaces were included in the study.

Of the 44 "standard" data sets, 12 contained hair, 11 had tilted surfaces and 2 contained both types of features. Of the 11 data sets considered "difficult", 2 contained hair, 5 had tilted surfaces and 2 contained both types of features.

Each volume had a size of 125×75×250 voxels with a definition of 40×40×8 μm.

1) Manual Ground Truth Generation

There is no established method to estimate the skin position in RSOM volumes. Therefore, we used manual tracings of the skin boundary as the gold standard to evaluate the performance of our method. For each volume V, three 2D image slices in the x-direction and three slices in the y-direction were randomly selected. In each image slice, the reference contour was generated to fit the skin contour by an experienced analyst. The manual task consisted of sequentially positioning control points to build a piecewise-interpolated polynomial.

2) Automatic Segmentation

The surface detection algorithm was applied to the RSOM volumes using a fixed set of empirically determined parameters. The neighborhood parameter $d = 3$ and the smoothness parameter $\delta = 0.2$ were fixed for all volumes, as the smoothness constraints for all surfaces in the datasets were the same. The parameters for the energy computation should be chosen based on the quality and the information content in the data. They can either be tuned for each data set to get the best possible output, or be fixed *a priori* based on a small set of representative training data. Since tuning parameters can be difficult and time-consuming, especially in clinics, we chose to focus on a fixed set of weights. For the given data, we empirically determined the weighting parameters (α = 3, β = 2, γ = 1) while giving the gradient energy the greatest importance and the noise energy the least. A standard deviation $\sigma = 3$ was defined for the Gaussian in the gradient energy computation.

In average cases, the weight α of the gradient energy, which emphasizes interfaces between anatomical structures, should be higher than the weights for the other energy components. When the skin boundary itself has low contrast, α should be increased. β weights the surface energy, which represents the probability that a voxel belongs to the uppermost surface. When prominent high-contrast structures can be found in deeper tissues, β should be increased. For its part, γ weights the noise energy, which prevents the algorithm from misclassifying surface-like artefacts as skin boundary. In cases with highly prominent surface-like noise structures, γ should be increased. When the contained surfaces are highly tilted, the noise energy does not contribute much and, therefore, can be set to 0. σ should be chosen depending on the extent of inconsistencies. Larger values of σ help to improve the surface detection even over larger inconsistencies, while at the same time smaller changes in the contour might get neglected.

3) Accuracy Evaluation

The performance of the proposed surface detection method was assessed by comparing it with manual segmentations. Each point $p_m^{(i)}$ on the manually traced contours was compared to the corresponding point $p_a^{(i)}$ in the automatically detected boundary surface with the same (x, y)-position using the mean absolute error (MAE) ± standard deviation and the maximum error.

An additional examination of the inter- and intra-analyst variability in the manual segmentation was performed to demonstrate the reliability of the ground truth used for the accuracy evaluation. Three analysts (A_1, A_2, A_3) – one with limited, one with intermediate and one with extensive experience in RSOM image analysis – independently segmented a predefined random subset of the dataset ($n = 10$). The analyst with limited experience, A1, had been working with optoacoustic data for only a short period and had never previously performed manual skin segmentation. The highly experienced analyst A3 had been analyzing optoacoustic datasets on a daily basis for several months and had regularly performed manual skin surface segmentations. All analysts were given the same information about the imaging modality, the visual appearance of skin in RSOM data and the task to perform. Additionally, all of them were instructed in the same way and were given the opportunity to test the manual segmentation on the same example data sets prior to the actual segmentation. The mean absolute error (MAE) ± standard deviation was computed from the maximum difference between each $p_{A_1}^{(i)}$, $p_{A_2}^{(i)}$ and $p_{A_3}^{(i)}$, and this MAE was used as a measure of inter-analyst variability in manual segmentation. To assess intra-analyst variability, the highly experienced analyst manually segmented the same dataset two more times at intervals of 24 h, and the same metrics were computed.

III. RESULTS

We tested our algorithm by applying it to RSOM datasets from the volar forearm of healthy volunteers (n = 55). Most of the datasets, referred here to as "standard cases" (n = 44), were expected to involve less challenging contour segmentation

because the contours were continuous and lay in regions of the image where the signal-to-noise ratio was relatively high. To test the robustness of the algorithm, we also used datasets termed "difficult cases" (n = 11), where we expected contour segmentation to be more difficult because of large contour discontinuity or poor contrast for the skin boundary, making these cases difficult even for manual segmentation. We applied our automated algorithm to both types of cases and then assessed the contour segmentation as "successful", "semi-successful" or "failed" based on qualitative visual analysis of the results (Table 1). If the algorithm detected the actual skin boundary, the segmentation was categorized as successful. If the detected surface followed the pathway of the skin boundary, but instead of the uppermost layer, a lower layer was detected, the result was categorized as semi-successful. In all other cases, results were categorized as failed.

Case	Hair	Tilt	n	Segmentation results*		
				Successful	Semi-successful	Failed
Standard			19	17	2	
	yes		12	9	3	
		yes	10	9	1	
	yes	yes	3	2		1
Difficult			2		2	
	yes		2	1	1	
		yes	5	2	1	2
	yes	yes	2			2
			55	40	10	5

* See text for evaluation criteria

Table 1. Performance of the automated contour segmentation algorithm with RSOM data from healthy volunteers.

Fig. 4 shows successful segmentation results of the proposed method on two data sets, both in the volume (Fig. 4 (a), (c)) and in a selected 2D slice (Fig. 4 (b), (d)) of the volume. While no smoothing was performed by the algorithm itself, we did smooth the surface purely for visualization purposes, after we had quantitatively analyzed its accuracy. The 2D slices clearly depict how the extracted boundaries cover the subcutaneous vasculature, on horizontal as well as tilted surfaces.

Overall, the detection algorithm successfully detected the skin surface in 40 of 55 volumes (~73%) examined, even though 12 of the 40 volumes contained local inaccuracies due to noise, artifacts resembling surface structures, or hair (8 volumes). The segmentation results were classified as semi-successful in 10 volumes. A comparison of the manual skin tracings and the automatic detection for the complete test data set, excluded the failed data sets, resulted in an MAE of 52.6 ± 26.6 μm (equivalent to an average offset of 7 ± 3 voxel) and an average maximum error of 122.3 ± 59.3 μm (equivalent to an offset of 15 ± 7 voxel). An MAE of 40.6 ± 13.5 μm was achieved in the successful segmentation cases. Fig. 5 (a) shows that most cases showing high error were also categorized as "difficult cases" during manual segmentation. At the same time, the algorithm managed to segment the surface in some of the difficult data sets at similar locations as the human analyst.
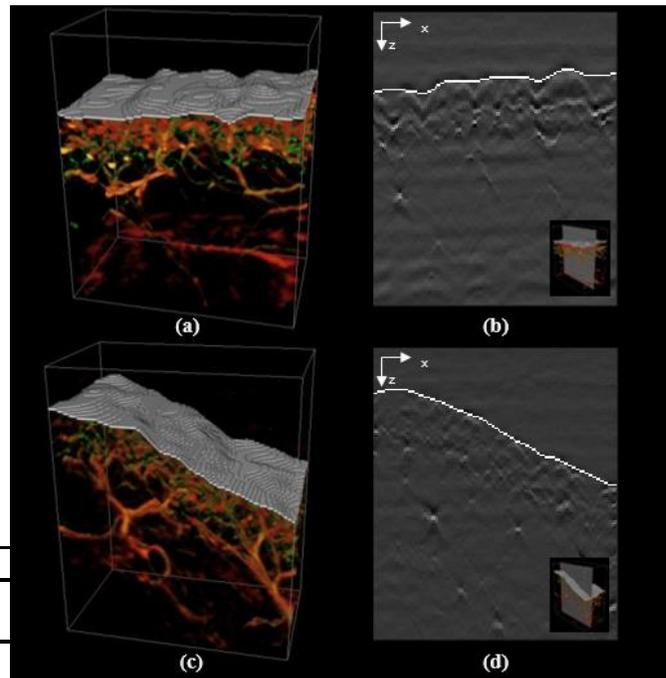

Fig. 4. Examples of successful segmentation of skin from RSOM data. These were standard cases without hair (first and third row of Table 1). **(a, c)** 3D volumes including the estimated skin layer. **(b, d)** 2D slices of the volumes in panels (a) and (c), respectively, including the estimated skin contour.

Table 2 compares the performance of the automatic segmentation with manual segmentation in light of inter- and intra-analyst variability during the manual process. The manual tracings of one analyst were similar over multiple executions, while differences between segmentations by different analysts were almost twice as large. The average difference of the automatic method was slightly greater than the inter-analyst variability.

Table 2. Inter- and intra-analyst variability*

	Difference (in μm)	
	Mean absolute error	Average maximum difference
Inter-analyst variability (A_1, A_2, A_3)	30.8 ± 8.0	88.9 ± 49.0
Intra-analyst variability (A_3)	15.8 ± 4.1	55.8 ± 24.7
Auto vs. Analysts**	49.4 ± 18.1	126.5 ± 103.2

* computed by comparing manual segmentations of 10 randomly selected data sets with the corresponding automatic segmentation
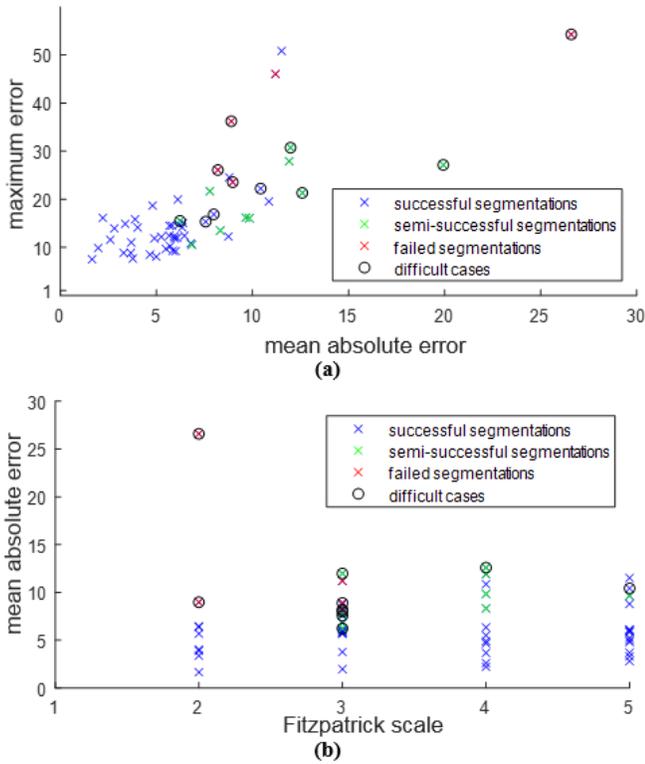** For 9 datasets after exclusion of one failed segmentation.

Fig. 5. Distribution of automatic segmentation results according to errors in voxels. Criteria for deciding whether segmentations were successful, semi-successful or failed are given in the text. "Difficult cases" (empty circles) refer to datasets containing large discontinuities and a highly diffuse surface layer, which made manual boundary detection extremely difficult. **(a)** Mean absolute error in comparison to the maximum error in each case. **(b)** Distribution of segmentation error in each Fitzpatrick phototype.

Fig. 6 (a) and (b) illustrate how hair can compromise the accuracy of the segmentation. While the algorithm successfully managed to separate the actual skin boundary from the hair perpendicular to the surface (Fig. 6 (d)), the hair running more parallel to the skin was classified as part of the surface (Fig. 6 (c)). Even though this segmentation met our criteria of success, it is clear that hair can compromise the detection accuracy. Hair lying close to the skin is particularly difficult to distinguish from the skin surface and is likely to cause inaccuracies in surface detection.

A frequent cause of semi-successful segmentation was when the uppermost layer showed diffuse contours, the skin boundary was predominantly discontinuous, and low-contrast subcutaneous structures were present. Fig. 6 (f) and (g) show that in such cases, the proposed method tended to classify the superficial subcutaneous structures as noise and identified the skin boundary as lying in a lower, less diffuse skin layer with increased contrast. Even during manual segmentation, such cases often required manual thresholding of the data in order to distinguish noise from subcutaneous structures.

The algorithm failed entirely to identify a skin boundary in 5 volumes, and the actual skin surface in all these cases was

highly tilted. In addition, most of these cases contained ill-defined skin layers and large discontinuities in the skin contours, leading them to be classified as "difficult cases" during manual segmentation.
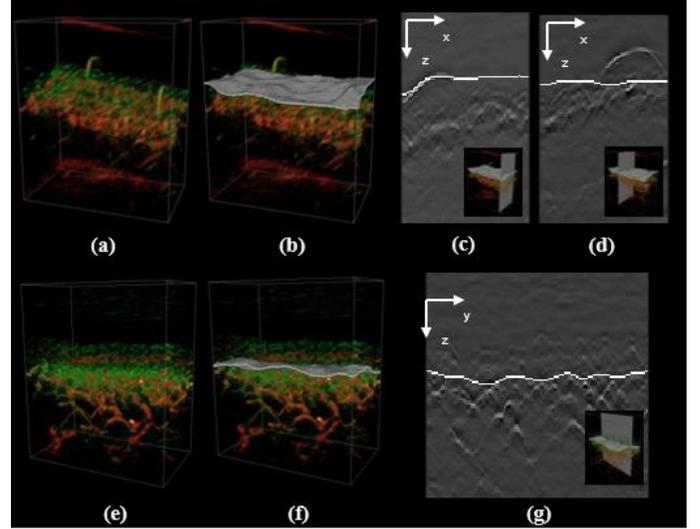

Fig. 6 Situations in which the automated algorithm did not segment contours well from RSOM data. Segmentation was inaccurate for volumes containing hair (a-d) and skin layers with low contrast (e-g). **(a, e)** 3D volumes. **(b, f)** The same 3D volumes as in panels (a) and (e), overlaid with the estimated skin layer. **(c, d, g)** 2D slices, with the estimated skin contour shown as a white line. The two 2D slices in the upper row (c, d) show the two different effects hair can have on the segmentation. The insets show the location of the 2D slices in the respective volume.

For cases with inaccuracies (n = 12), we additionally examined whether tuning the initial parameters based on the data quality and content (as explained in section II.F.2) would improve the segmentation accuracy. The initial segmentation was visually assessed, then manually tuned where judged necessary. This manual refinement improved the visual results in 9 datasets[1]. Manual tuning eliminated local inaccuracies from 6 volumes and reduced them in the other 3 volumes. The MAE of successful segmentations was reduced to 38.8 ± 14.4 μm.

The Fitzpatrick skin phototype and consequently the skin complexion and the melanin content in the skin did not seem to play a significant role in the segmentation process, at least for the phototypes investigated in this study. The distribution of successful and failed segmentations was similar across the different Fitzpatrick values. Fig. 5 (b) demonstrates the distribution of segmentation error over the different phototypes. A pairwise comparison of the error distribution in each Fitzpatrick scale with Student's t-test upholds the null hypothesis (with p-values > 0.3) and, thus, shows that skin complexion did not have a significant impact on the segmentation accuracy.

An important characteristic of our method is its direction-independency, since it considers all possible solutions (ie. the candidate z-coordinate of the surface during its propagation

---

[1] See also Supplementary Table 1 for a more detailed breakdown of the segmentation accuracy when using tuned parameters as opposed to fixed ones.

Supplementary materials are available in the supplementary files.

along both x and y directions), without taking any final decision before the full volume has been interrogated. In contrast, the Cheng and Lin method [22] first calculates an intermediate map based on propagation along one direction only (x or y) and then uses this map as *a priori* information to generate the final map (namely, by propagating along the other spatial direction). This two-step method may provide different results depending on the order of propagation directions. In our approach, however, a single propagation takes place simultaneously along both spatial directions, so information along x and y is simultaneously considered in order to solve the problem. The final decision of determining the optimal surface is therefore taken only after the full volume has been assessed, thereby providing a unique solution.[2]

IV. DISCUSSION

In this work, a method to extract the skin surface in RSOM volumes was presented. It follows the DP strategy and detects the surface in one step using front propagation with subsequent backtracking. This work extends previous studies on contour segmentation by introducing a way to perform front propagation in two dimensions simultaneously (as opposed to one after the other) and, therefore, segmenting a surface instead of multiple contours. Thus, our method is direction-independent and yields a unique result, as opposed to methods performing front propagation in one dimension after the other [22], which may generate different segmentations depending on the order of directions. Furthermore, it provides a way to extract the skin boundary depicted in optoacoustic mesoscopy by presenting a domain-specific energy function that takes the properties of the imaging modality into account, such as surface-like artifacts. Replacing the RSOM-specific energy function with an appropriate application-specific energy function may allow our approach to be applied to any surface detection task in numerous biomedical imaging contexts and with other imaging modalities.

The proposed method was evaluated against 2D manual annotations of *in vivo* measurements. While no real ground truth is available for such measurements and the manual tracings that serve as the basis for the evaluation are subjective, they give the best overview of the capabilities and limitations of the method. Simulated phantoms possess a ground truth but depict only a highly simplified scenario, whereas physical phantoms show a more realistic scenario, but lack ground truth due to deformations during the scanning process. Therefore, despite the lack of a real ground truth, we chose *in vivo* measurements with manual tracings for evaluation to assess the clinical translatability of our method. These manual reference tracings, however, are predisposed to error and inter-observer variability. In particular, the location of the skin boundary often can only be estimated based on the location of subcutaneous vasculature; as a result, different users may segment the same

surface differently, and this can even be true of the same user segmenting the surface at different points in time. This is particularly true for regions with discontinuities. During manual segmentation of these areas, the contour is estimated by interpolating the information contained in the neighboring regions along one dimension, whereas information along the other dimension is neglected, which can cause inaccuracies. The proposed method, in contrast, extracts the skin surface by taking the information contained in the whole volume into account. It always delivers the same result for the same volume and same parameter values.

Furthermore, taking the axial resolution into account, the MAE of 53 ± 27 μm is equivalent to a segmentation offset of around 7 ± 3 voxels with respect to manual segmentation. This error for automatic segmentation is small compared to human error and variability of nearly 4 ± 1 voxels in the reference segmentations.

Even though the proposed 3D boundary detection method was developed for the extraction of the skin surface in RSOM volumes, 2D front propagation can be used for various open surface detection tasks in different image modalities. The general method is universal; solely the computed volume energy is RSOM- and task-specific. By changing the energy function according to the task in hand, the method can be applied to other surface detection problems as, for instance, the segmentation of arteries in volumetric intravascular ultrasound images in polar domain [27, 28].

The surface extraction is fast and efficient. Once the volume energy is computed, the processing time for the extraction itself is only linearly dependent on the volume size and the chosen neighborhood parameter d with an overall complexity of $O(mnod)$. In the present study, the surface detection took around 2.3 sec per volume, with the energy computation consuming up to 75% of the time[3]. Our method, hence, has faster processing times and requires less human interaction than manual tracing, which can take 2-5 min even when performed on only a few volume slices. Previous methods, such as the algorithm presented by Cheng and Lin [22] take up to 5 seconds (after adaptation to RSOM segmentation), as they require two propagations through the volume.

Unlike previous approaches, the segmentation is not performed slice-wise or even per A-line with subsequent smoothing or interpolation. Instead, information contained in the entire volume simultaneously contributes to the surface detection. Subsequently, the boundary surface is extracted in one pass, without requiring any smoothing or interpolation of piecewise extracted results.

While in our study it was not needed, the algorithm in general allows the user to constrain the smoothness of the extracted surface by choosing δ accordingly, if required. However, enforcing increased smoothness results in lower accuracy of the segmentation. The better the energy function is designed for a given task, the better the detection works. Hence, the choice of

---

[2] See Supplementary Fig. 1 for a comparison of the direction-(in)dependency of both methods. Supplementary materials are available in the supplementary files.

[3] See Supplementary Table 2 for a comparison of execution times for different processing steps and different segmentation methods and their substeps. Supplementary materials are available in the supplementary files.

the initial parameters α, β and γ is critical for the outcome of the segmentation and should be done based on the task in hand and content and quality of the overall data. In this sense, the proposed surface extraction method is not fully automatic but semi-automatic for the case of skin surface detection in RSOM, as it requires selection of the weighting parameters for the energy computation based on the task at hand. Furthermore, data quality is an essential criterion for the success of the segmentation. Poor image quality, increased noise level and, especially, a poor-quality skin boundary can obstruct the extraction process and cause inaccuracies or even failure in extreme cases.

In conclusion, the location of the skin surface is essential information for automatic analysis of dermatological biomarkers in RSOM volumes. A 3D surface segmentation method to extract the skin boundary is proposed. Based on dynamic programming, the approach localizes the skin boundary even over discontinuities and diffuse interfaces, producing faster and more reproducible results than manual segmentation. Hence, this method represents the first step towards a more automated approach to analyze biomarkers in skin tissue and deliver diagnoses, which in the long term can help establish RSOM as a diagnostic tool in clinical settings.

V. ACKNOWLEDGMENT

REFERENCES

[1] C. E. Griffiths and J. N. Barker, "Pathogenesis and clinical features of psoriasis," *Lancet*, vol. 370, no. 9583, pp. 263–271, 2007.

[2] W. H. Clark, L. From, E. A. Bernardino, and M. C. Mihm, "The histogenesis and biologic behavior of primary human malignant melanomas of the skin.," *Cancer Res.*, vol. 29, no. 3, pp. 705–27, Mar. 1969.

[3] S. P. Barton, M. S. Abdullah, and R. Marks, "Quantification of microvascular changes in the skin in patients with psoriasis," *Br J Dermatol*, vol. 126, no. 6, pp. 569–574, Jun. 1992.

[4] A. Srivastava, P. Laidler, L. E. Hughes, J. Woodcock, and E. J. Shedden, "Neovascularization in human cutaneous melanoma: A quantitative morphological and Doppler ultrasound study," *Eur. J. Cancer Clin. Oncol.*, vol. 22, no. 10, pp. 1205–1209, 1986.

[5] J. Aguirre, M. Schwarz, N. Garzorz, M. Omar, A. Buehler, K. Eyerich, and V. Ntziachristos, "Precision assessment of label-free psoriasis biomarkers with ultra-broadband optoacoustic mesoscopy," *Nat. Biomed. Eng.*, vol. 1, no. 5, p. 0068, 2017.

[6] R. Joannides, J. Bellien, and C. Thuillez, "Clinical methods for the evaluation of endothelial function - A focus on resistance arteries," *Fundam. Clin. Pharmacol.*, vol. 20, no. 3, pp. 311–320, 2006.

[7] J. C. Patik, K. M. Christmas, C. Hurr, and R. M. Brothers, "Impaired endothelium independent vasodilation in the cutaneous microvasculature of young obese adults," *Microvasc. Res.*, vol. 104, pp. 63–68, Mar. 2016.

[8] L. Guilherme Kraemer-Aguiar, C. Maurente Laflor, and E. Bouskela, "Skin microcirculatory dysfunction is already present in normoglycemic subjects with metabolic syndrome," 2008.

[9] J. Cui, A. Arbab-Zadeh, A. Prasad, S. Durand, B. D. Levine, and C. G. Crandall, "Effects of Heat Stress on Thermoregulatory Responses in Congestive Heart Failure Patients," *Circulation*, vol. 112, no. 15, pp. 2286--2292, 2005.

[10] V. Ntziachristos, "Going deeper than microscopy: the optical imaging frontier in biology," *Nat. Methods*, vol. 7, no. 8, pp. 603–614, Aug. 2010.

[11] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography.," *Science*, vol. 254, no. 5035, pp. 1178–81, Nov. 1991.

[12] S. Jiao, Z. Xie, H. F. Zhang, and C. A. Puliafito, "Simultaneous multimodal imaging with integrated photoacoustic microscopy and optical coherence tomography.," *Opt. Lett.*, vol. 34, no. 19, pp. 2961–3, Oct. 2009.

[13] G. Liu, W. Jia, V. Sun, B. Choi, and Z. Chen, "High-resolution imaging of microvasculature in human skin in-vivo with optical coherence tomography.," *Opt. Express*, vol. 20, no. 7, pp. 7694–705, Mar. 2012.

[14] P. Beard, "Biomedical photoacoustic imaging," *Interface Focus*, vol. 1, no. 4, pp. 602–631, 2011.

[15] M. Omar, J. Gateau, and V. Ntziachristos, "Raster-scan optoacoustic mesoscopy in the 25–125 MHz range," *Opt. Lett.*, vol. 38, no. 14, pp. 2472–2474, 2013.

[16] M. Omar, D. Soliman, J. Gateau, and V. Ntziachristos, "Ultrawideband reflection-mode optoacoustic mesoscopy," *Photoacoust. imaging*, vol. 1705120, no. 110, 2014.

[17] J. Aguirre, M. Schwarz, D. Soliman, A. Buehler, M. Omar, and V. Ntziachristos, "Broadband mesoscopic optoacoustic tomography reveals skin layers," *Opt. Lett.*, vol. 39, no. 21, p. 6297, 2014.

[18] J. Aguirre, B. Hindelang, A. Berezhnoi, U. Darsow, F. Lauffer, K. Eyerich, T. Biedermann, and V. Ntziachristos, "Assessing nailfold microvascular structure with ultra-wideband raster-scan optoacoustic mesoscopy," *Photoacoustics*, vol. 10, pp. 31–37, Jun. 2018.

[19] M. Omar, M. Schwarz, D. Soliman, P. Symvoulidis, and V. Ntziachristos, "Pushing the Optical Imaging Limits of Cancer with Multi-Frequency-Band Raster-Scan Optoacoustic Mesoscopy (RSOM)," *NEO*, vol. 17, pp. 208–214, 2015.

[20] H. F. Zhang, K. Maslov, and L. V. Wang, "Automatic algorithm for skin profile detection in photoacoustic microscopy," *J. Biomed. Opt.*, vol. 14, no. 2, p. 024050, 2009.

[21] G. Zahnd, A. Karanasos, G. van Soest, E. Regar, W. Niessen, F. Gijsen, and T. van Walsum, "Quantification of fibrous cap thickness in intracoronary optical coherence tomography with a contour segmentation method based on dynamic programming," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 9, pp. 1383–1394, Sep. 2015.

[22] D.-C. Cheng and J.-T. Lin, "Three-dimensional expansion of a dynamic programming method for boundary detection and its application to sequential magnetic resonance imaging (MRI).," *Sensors (Basel).*, vol. 12, no. 5, pp. 5195–211, 2012.

[23] S. Astner and R. R. Anderson, "Skin Phototypes 2003," *J. Invest. Dermatol.*, vol. 122, no. 2, pp. xxx–xxxi, 2004.

[24] K. Ungru and X. Jiang, "Dynamic Programming Based Segmentation in Biomedical Imaging," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 255–264, 2017.

[25] R. Bellman, "Dynamic programming.," *Science*, vol. 153, no. 3731, pp. 34–7, Jul. 1966.

[26] S. R. Eddy, "What is dynamic programming?," *Nat. Biotechnol.*, vol. 22, no. 7, pp. 909–910, Jul. 2004.

[27] R. I. Kitney, C. Burrell, L. Moura, K. Straughan, and M. T. Rothman, "3D characterization of the arterial wall using intravascular ultrasound," in *[1991] Proceedings Computers in Cardiology*, pp. 357–360.

[28] D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation," 2000.