Segmentation of Subcutaneous Fat within Mouse Skin in 3D OCT Image Data using Random Forests

Timo Kepp\textsuperscript{a}, Christine Droigk\textsuperscript{c}, Malte Casper\textsuperscript{b,c}, Michael Evers\textsuperscript{b,c}, Nunciada Salma\textsuperscript{c}, Dieter Manstein\textsuperscript{c}, and Heinz Handels\textsuperscript{a}

\textsuperscript{a}Institute of Medical Informatics, University of Lübeck, Germany
\textsuperscript{b}Institute of Biomedical Optics, University of Lübeck, Germany
\textsuperscript{c}Cutaneous Biology Research Center, Massachusetts General Hospital, Boston, MA

ABSTRACT

Cryolipolysis is a well-established cosmetic procedure for non-invasive local fat reduction. This technique selectively destroys subcutaneous fat cells using controlled cooling. Thickness measurements of subcutaneous fat were conducted using a mouse model. For detailed examination of mouse skin optical coherence tomography (OCT) was performed, which is a non-invasive imaging modality. Due to a high number of image slices manual delineation is not feasible. Therefore, automatic segmentation algorithms are required. In this work an algorithm for the automatic 3D segmentation of the subcutaneous fat layer is presented, which is based on a random forest classification followed by a graph-based refinement step. Our approach is able to accurately segment the subcutaneous fat layer with an overall average symmetric surface distance of $11.80 \pm 6.05 \mu m$ and Dice coefficient of $0.921 \pm 0.045$. Furthermore, it was shown that the graph-based refining step leads to increased accuracy and robustness of the segmentation results of the random forest classifier.

Keywords: optical coherence tomography, random forest classification, graph-based segmentation

1. PURPOSE

Unwanted excessive fat is one of the main reasons for a cosmetic surgery. Furthermore, media promotes a slim body shape. Cryolipolysis is a novel non-invasive method for cold-induced selective destruction of subcutaneous fat tissue.\textsuperscript{1} This results in a reduction of the subcutaneous fat layer (SFL). Cryolipolysis was applied over four million times in humans with numerous clinical studies.\textsuperscript{2} In a first study the efficacy of the cryolipolysis technique was determined using a pig model.\textsuperscript{3} Ultrasound imaging was performed in order to measure the thickness of the fat layer before and after the treatment, while the position with the largest reduction was measured. Since pig and clinical studies are very complex and cost-intensive, there is a need for a new approach using a mouse model. Therefore, we acquired images of the SFL of mouse skin with optical coherence tomography (OCT), which is an imaging technique based on the scattering properties of near infrared light in tissue structures.\textsuperscript{4} This non-invasive modality provides three-dimensional images in micrometer-scale resolution and is frequently used in the field of ophthalmology. Also in other clinical areas such as dermatology OCT is gaining popularity enabling the evaluation of skin lesions and melanoma. In this paper OCT imaging technique is used to detect the SFL thickness of mice. The SFL is segmented in OCT volumes, which has been taken of the inguinal region of the mouse. Some previous works with a similar focus has been proposed. In their work Sheet et al.\textsuperscript{5} characterized OCT images of the mouse skin and compared them with corresponding histologies. They used a transfer learning approach to incorporate statistical physics models into a random forest (RF) classifier.

In our previous work RF classifier demonstrated high accuracy in various medical image data, such as for the segmentation of ischemic stroke lesions in MR image data.\textsuperscript{6, 7} Hence, we use an RF classifier for the automatic segmentation of multiple tissue layers within mouse skin in 3D OCT image data in this paper. RF classifier are very popular because of their flexibility, few parameters and ease of use. In addition, the RF classifier naturally handles multi-class problems. Instead of a sophisticated transfer learning function\textsuperscript{8} we combine the RF classifier with a graph-based approach that ensures spatial consistency and is computationally efficient.

Send correspondences to Timo Kepp (kepp@imi.uni-luebeck.de)
2. METHODS

The main goal of our work is to classify and segment the subcutaneous fat layer (SFL) within mouse skin in 3D OCT image data. Since the SFL is located between the dermis layer (DL) and muscle or fascial layer (ML/FL), we integrated these as boundary features into the classification problem for improved robustness. Those additional layers will be used to improve the segmentation result during the refinement step. To ensure that the same area is imaged over a period of time it was marked using micro-tattoos. These tattoos are also integrated into the classification process. Unfortunately, the epidermis layer is hard to distinguish at contact points on the sample spacer (see Figure 1a). We therefore treat epidermis and dermis as one layer. Thus, we distinguish between six classes: background, DL, SFL, FL, ML and Tattoo. Figure 1a shows an example of a cross-sectional view (B-scan) through a OCT volume. In the following preprocessing (Sec. 2.1), RF classification (Sec. 2.2) and the graph-based refinement (Sec. 2.3) are described in detail. A flowchart of our algorithm is shown in Figure 2.

2.1 Preprocessing

Initially, we perform several preprocessing steps to reduce noise and to transform each image into a common space. Due to the very large image size we reduce the image resolution by half. In their raw version OCT image data are very noisy (speckle noise). Since the noise would greatly affect the computation of the RF features, we use the BM3D filter\(^8\) to overcome this issue. Furthermore, we compensate local brightness differences, which can occur within a B-scan and normalize the intensity values. We then flatten all OCT volumes to reduce the variability of the RF features, which could lead to worse classification results. Here, we use the dermis profile as a reference structure, which is detected roughly by using an average filter in combination with thresholding.
A second order polynomial is then fitted through the estimated dermis profile. To flatten the image we shift each column of the image so that the fitted polynomial results in a straight horizontal line. This procedure is performed on every 20th B-scan and shifts of all remaining B-scans were linear interpolated. As a last step we crop all image volumes to remove unnecessary background areas and to reduce computation time. An example B-scan before and after the preprocessing is shown in Figure 3.
Table 1: List of used features and their entries in the feature vector.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Number of entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity</td>
<td>1</td>
</tr>
<tr>
<td>Cumulative sum:</td>
<td></td>
</tr>
<tr>
<td>- Column-wise, above and below the current position</td>
<td>2</td>
</tr>
<tr>
<td>- Column-wise, mean value over row</td>
<td>1</td>
</tr>
<tr>
<td>- Local mean value for each row of the patch</td>
<td>15</td>
</tr>
<tr>
<td>Local standard deviation</td>
<td>1</td>
</tr>
<tr>
<td>Local intensity histogram (5 bins: 0, 0.2, 0.4, 0.6, 0.8, 1)</td>
<td>5</td>
</tr>
<tr>
<td>Difference of Gaussian:</td>
<td></td>
</tr>
<tr>
<td>- Filter response</td>
<td>1</td>
</tr>
<tr>
<td>- Mean value of filter response in $15 \times 15$ neighborhood</td>
<td>1</td>
</tr>
<tr>
<td>Binary sobel</td>
<td>1</td>
</tr>
<tr>
<td>Vertical edges</td>
<td>1</td>
</tr>
<tr>
<td>Horizontal edges</td>
<td>1</td>
</tr>
<tr>
<td>Distance to closest edge &amp; relative position</td>
<td>3</td>
</tr>
<tr>
<td>Sum of local gradient magnitude</td>
<td>1</td>
</tr>
<tr>
<td>BRIEF</td>
<td>30</td>
</tr>
<tr>
<td>Census</td>
<td>30</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>94 \times 3 \text{ level } = 282</strong></td>
</tr>
</tbody>
</table>

### 2.2 Random forest classification

In our approach we use an RF classifier to segment six different classes. RF classifiers were first described by Breiman et al. and consist of an ensemble of decision trees.\(^9\) In our case the RF classifier consists of $T = 60$ decision trees. The RF classifier uses features to decide how voxels are classified. Some features, such as image intensity or edge filter response, can be used for almost any classification problem. However, most of the features have to be designed specifically for the problem. We utilize voxel-based and patch-based features in our approach.

Some features are based on the image intensity and are calculated both voxel-wise and patch-based (raw intensity, cumulative sums, local standard deviation, local histogram). Furthermore, edge information is used. To learn the honeycomb structure of the SFL, we use the Difference of Gaussian. To integrate more context information to the feature space BRIEF descriptor and census transform\(^10\) are used for feature extraction. Eventually, each feature is computed at three different scales using a Gaussian pyramid. This results in a feature vector with a total of 282 components (Tab. 1). In order to ensure randomness we use both bagging and random feature selection. For bagging we use $5D$ random data points for the training of each tree and for each split of a node we optimize over $M_k = \sqrt{M}$ features. Here, $D$ describes the number of data and $M$ represents the number of features. Since adjacent B-scans are very similar, we have trained the RF classifier on 2D B-scan images.

During the training process the classifier uses label data, which represent the ground truth. We train the RF classifier with features extracted from training images using a sparse uniform grid. Once trained, the classifier can be applied to unlabeled image data by extracting the same features and put them into the classifier to obtain label probabilities. For computation efficacy we use the sparse grid for a first classification. Voxels, which have not been classified, are interpolated using a nearest neighbor interpolation. We then perform a second voxel-wise classification at potential layer borders of the segmentation results to achieve a higher accuracy. To eliminate outliers in the classification results we apply region growing and closing operations as postprocessing steps.

### 2.3 Graph-based refinement

To take neighborhood information during voxel-wise classification into account an additional refinement step integrates neighboring B-scans and expands the segmentation approach to the three-dimensional space. Here, the segmentation by the RF classification is used as a shape prior. This refinement step is based on an optimal surface detection using graph-theory. Furthermore, this introduces a global regularization due to the conversion of voxel-wise classification into a multi-surface-based detection and ensures spatial consistency at the same time.
The framework described here is based on the work of Li et al.\cite{li2011graph} and has been adjusted for our purpose. Lang et al.\cite{lang2006boundary} use the same framework for boundary optimization after RF classification. Antony et al.\cite{antony2005simplifying} use variation for the refinement of several RF classifications.

The general basis is the construction of a node-weighted digraph $G = (V, E)$ with nodes $V$ and arcs $E$. Here, each arc $(v_i, v_j) \in E$ connects nodes $v_i, v_j \in V$. The node set $V$ of the graph consists out of all image voxels, where each node belongs to a single voxel $(x, y, z) \in \Omega^{(3)}$ denoted as $v_{x,y,z}$. Thus, $V = \{v_{x,y,z}| (x, y, z) \in \Omega^{(3)}\}$. The cost function $c: \Omega^{(3)} \rightarrow \mathbb{R}$ assigns a cost value to each voxel. The costs $\tilde{w}(v_{x,y,z})$ calculated as follows:

$$
\tilde{w}(v_{x,y,z}) := \begin{cases} 
c(x, y, z), & \text{if } z = 0 \\
c(x, y, z) - c(x, y, z - 1), & \text{otherwise.} 
\end{cases}
$$

The computation of the optimal surface on the graph is equivalent to find the minimum closed set, which can efficiently be determined by using minimum s-t cuts. A further description of the graph construction can be found in the work of Li et al.\cite{li2011graph}

The cost function design is crucial for the segmentation quality and is described in the following. In general, cost functions are based on image edges or intensities in the local environment. We construct a cost function based on the image gradients and the post-processed RF classification. For each of the two surfaces an individual cost function has to be determined. The surfaces describe the transition from DL to SFL and from SFL to FL/ML, respectively. To integrate edge information into the cost function we compute at each position $(x, y, z) \in \Omega^{(3)}$ the gradient magnitude, which we normalize and invert to obtain low costs at voxel positions with high gradient magnitudes:

$$
c_{\text{grad}}(x, y, z) = 1 - \frac{|\nabla I(x, y, z)|}{\max_{\tilde{x}, \tilde{y}, \tilde{z} \in \Omega^{(3)}} |\nabla I(\tilde{x}, \tilde{y}, \tilde{z})|}.
$$

Next, the cost of the DL-SFL transition is described. On the one hand distances to the DL and on the other hand distances to the SFL are used for the construction of the cost function. Afterwards we filter the binarized RF segmentation of the DL and SFL with a vertical edge filter. Thereby, we receive negative and positive filter responses at the upper and lower segmentation boundaries. Here, $D^-$ and $F^+$ describe sets of voxel for which the filter response on the DL is negative and positive on the SFL, respectively. By adding up the minimal distances of those locations, we receive the following cost function for the surface between DL and SFL:

$$
c^1(x, y, z) = \min_{(\tilde{x}, \tilde{y}, \tilde{z}) \in D^-} ||(x, y, z) - (\tilde{x}, \tilde{y}, \tilde{z})||_2 + \min_{(\tilde{x}, \tilde{y}, \tilde{z}) \in F^+} ||(x, y, z) - (\tilde{x}, \tilde{y}, \tilde{z})||_2 + \omega_G \cdot c_{\text{grad}}(x, y, z),
$$

where $\omega_G$ is used as a scaling factor for $c_{\text{grad}}$. The second surface delineates the SFL from the FL/ML. Since the FL is difficult to segment manually and also a direct transition from SFL to ML is possible, we consider FL and ML as a single class. The cost function construction is analog:

$$
c^2(x, y, z) = \min_{(\tilde{x}, \tilde{y}, \tilde{z}) \in F^-} ||(x, y, z) - (\tilde{x}, \tilde{y}, \tilde{z})||_2 + \min_{(\tilde{x}, \tilde{y}, \tilde{z}) \in F^+} ||(x, y, z) - (\tilde{x}, \tilde{y}, \tilde{z})||_2 + \omega_G \cdot c_{\text{grad}}(x, y, z).
$$

Referring to Garvin et al.\cite{garvin2012variational} we extend our cost function by in-region costs. Let $L_{RF}(x, y, z)$ be the class, which assigns the post-processed RF classification to voxel $(x, y, z)$. Both surfaces divide the image into three regions: background/DL (region 1), SFL (region 2), FL/ML/background (region 3). The in-region costs for both surfaces $i = 1, 2$ are then

$$
w^i_{\text{in-reg}}(v_{x,y,z}) := \begin{cases} 
w_R, & \text{if } L_{RF}(x, y, z) = \text{region } i \\
-w_R, & \text{if } L_{RF}(x, y, z) = \text{region } i + 1 \\
0, & \text{otherwise.} 
\end{cases}
$$

These costs are negative for the region $i + 1$, which should be below surface $i$ and positive for the overlying region $i$. Since the minimum closed set contains all nodes which are below the target surface, the surface will
preferably proceed above region \( i + 1 \) due to low costs and below region \( i \) showing high costs. The total costs \( w^i \) result from the addition of \( w^i_{in-reg} \) and equation 1:

\[
    w^i(x, y, z) := \begin{cases} 
    -1, & \text{if } z = 0 \\
    c^i(x, y, z) - c^i(x, y, z - 1) + w^i_{in-reg}, & \text{otherwise.}
\end{cases}
\] (6)

Note that the cost for \( z = 0 \) is changed to \(-1\) to prevent an empty minimum closed set.

For our purpose, the approach of Li et al. is modified as described in the following. Due to the fact that the mice were marked by tattoos, the tissue layers are pierced at these spots. Therefore, discontinuities are integrated into the graph model. Furthermore, the tattoo borders appear almost vertical (see Fig. 1a). Due to this second change, surfaces can only deform stronger nearby a tattoo.

### 3. EXPERIMENTS AND RESULTS

Our evaluation is based on 17 3D OCT volumes of subcutaneous fat in the inguinal region from 4 C57BL/6 mice before and after the cryolipolysis treatment. Scans were acquired on a Telesto II OCT system (Thorlabs, Inc.) with a FOV of 8.5 × 8.5 × 3.5 mm (spatial resolution 1307 × 1307 × 1024 voxel and spacing 6.5 × 6.5 × 3.4 μm). Note that the preprocessing steps (see Sec. 2.1) significantly reduce the original image size. After preprocessing, the spatial resolution is 653 × 653 × 200 (spacing 13 × 13 × 6.8 μm). Four individual B-scans per OCT volume were randomly selected and segmented by a human expert, which represent the ground truth. This results in a total number of 68 ground truth labeled B-scans. Our goal is to compare the segmentation results of the sole RF-based segmentation with the refined one by our graph-based approach (RF+GR). A leave-one-out scheme is performed to evaluate the segmentation performance. The segmentation accuracy was described with respect to the ground truth by computing Dice coefficient and average symmetric surface distance (ASSD).

![Dice](a) ![ASSD](b)

Figure 4: Box plots of the (a) Dice coefficients and (b) ASSDs.

Both Dice coefficient and ASSD results are visualized as box plots in Figure 4. On average the sole RF classifier provides good segmentation results with an ASSD of 17.40 ± 12.69 μm and Dice coefficient of 0.914 ± 0.047. Note that all results are averaged over all subjects. Furthermore, the plot in Figure 4b indicates that RF+GR improves the ASSD by 32 % (ASSD: 11.80 ± 6.05 μm, Dice: 0.921 ± 0.045). In addition, the standard deviation of the ASSD has decreased by 52 %, which is associated with an increased robustness. Moreover, a reduced number of
outliers can be detected. While the improved results of the ASSD are statistically significant (Wilcoxon rank-sum test, $p < 0.05$), unfortunately no significant improvement of the Dice coefficient can be observed. A qualitative comparison of the segmentation results of both methods is shown in Figure 5.

4. CONCLUSION

In this contribution we present a segmentation framework for the detailed evaluation of the DL and SFL thickness within mouse skin. An RF-based classification of the SFL with an additional graph-based refinement step is performed. The results demonstrate that our approach is capable of segmenting the target structures with high accuracy. Furthermore, we found the refining step to have a positive effect on the segmentation results and to guarantee higher accuracy and robustness.

REFERENCES


