

# Journal of Biomedical Optics

[SPIDigitalLibrary.org/jbo](http://SPIDigitalLibrary.org/jbo)

## **Pilot study of semiautomated localization of the dermal/epidermal junction in reflectance confocal microscopy images of skin**

Sila Kurugol  
Jennifer G. Dy  
Dana H. Brooks  
Milind Rajadhyaksha

# Pilot study of semiautomated localization of the dermal/epidermal junction in reflectance confocal microscopy images of skin

Sila Kurugol,<sup>a</sup> Jennifer G. Dy,<sup>a,\*</sup> Dana H. Brooks,<sup>a,\*</sup> and Milind Rajadhyaksha<sup>b,\*</sup>

<sup>a</sup>Northeastern University, Electrical and Computer Engineering, 360 Huntington Avenue, Boston, Massachusetts 02115

<sup>b</sup>Memorial Sloan-Kettering Cancer Center, Dermatology Service, 160 East 53rd Street, New York, New York 10022

**Abstract.** Reflectance confocal microscopy (RCM) continues to be translated toward the detection of skin cancers *in vivo*. Automated image analysis may help clinicians and accelerate clinical acceptance of RCM. For screening and diagnosis of cancer, the dermal/epidermal junction (DEJ), at which melanomas and basal cell carcinomas originate, is an important feature in skin. In RCM images, the DEJ is marked by optically subtle changes and features and is difficult to detect purely by visual examination. Challenges for automation of DEJ detection include heterogeneity of skin tissue, high inter-, intra-subject variability, and low optical contrast. To cope with these challenges, we propose a semiautomated hybrid sequence segmentation/classification algorithm that partitions z-stacks of tiles into homogeneous segments by fitting a model of skin layer dynamics and then classifies tile segments as epidermis, dermis, or transitional DEJ region using texture features. We evaluate two different training scenarios: 1. training and testing on portions of the same stack; 2. training on one labeled stack and testing on one from a different subject with similar skin type. Initial results demonstrate the detectability of the DEJ in both scenarios with epidermis/dermis misclassification rates smaller than 10% and average distance from the expert labeled boundaries around 8.5  $\mu\text{m}$ . © 2011 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.3549740]

Keywords: reflectance confocal microscopy; skin; dermal epidermal junction; texture segmentation; classification.

Paper 10373RR received Jul. 1, 2010; revised manuscript received Jan. 7, 2011; accepted for publication Jan. 10, 2011; published online Mar. 16, 2011.

## 1 Introduction

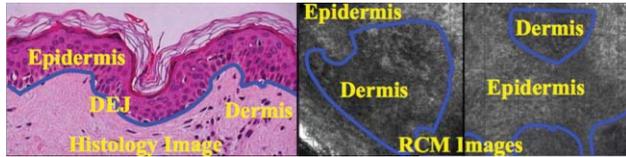
Skin cancer is among the most common forms of cancer in the USA, Australia, and Europe and is increasing in incidence in other parts of the world.<sup>1</sup> Visual inspection leading to biopsy followed by histology is the conventional method for clinical screening and diagnosis. Reflectance confocal microscopy (RCM) has been under development<sup>2</sup> for noninvasive imaging of skin for cancer screening and diagnosis.<sup>3-7</sup> RCM enables imaging and visualization of the epidermis and superficial dermis layers below the surface of the skin. Maximum imaging depth is limited to the papillary dermis or superficial reticular dermis, depending on the state of the overlying epidermis and the dermis/epidermis junction. Nuclear and cellular detail is imaged with nominal (instrumental) optical sectioning of 1 to 3  $\mu\text{m}$  and lateral resolution of 0.5 to 1.0  $\mu\text{m}$ , which is comparable to that of conventional pathology. RCM is advancing toward clinical application in dermatology for detection of malignancies such as melanoma and basal cell carcinomas with high sensitivity and specificity.<sup>3-6,8</sup> In these studies, sensitivity for detecting melanomas was reported to be 91% and specificity was 69%. Dermoscopy provides sensitivity of 80 to 100% and specificity of 70 to 100% but mainly for pigmented lesions.<sup>9,10</sup> Lightly pigmented, nonpigmented (amelanotic), and pink colored lesions are difficult to differentiate with dermoscopy.<sup>11</sup>

When examining all such types of lesions, in a more generalized setting, dermoscopy performs with specificity of 32 to 39%, compared to 68 to 84% with reflectance confocal microscopy.<sup>5</sup> For melanoma, while the sensitivity of RCM was comparable to that of the current standard of visual dermoscopy, the specificity was two times higher. Lentigo maligna melanomas were detected with sensitivity of 85% and specificity of 76%.<sup>6</sup> For detecting basal cell carcinomas, the sensitivity was 94% and specificity 78%.<sup>3</sup> These translational advances represent significant advances for RCM technology toward clinical utility. However, unlike histological sections, RCM images are oriented en face and are grayscale (unstained), and they do not visually resemble conventional stained pathology sections. Thus, differentiation of certain cellular features remains challenging. One such example of a clinically important feature is the dermis/epidermis junction.

Moreover, routine clinical use will require substantial training for clinicians. To accelerate such routine use and increase clinical acceptance, computer-automated methods to assist detection of morphologic features for screening and diagnosis may significantly expand the utility of RCM. Such methods may be either fully automated or semiautomated with user-assistance. However, as we describe below, this is a challenging application for automation with either method. One attempt toward automation was reported by Koller et al.<sup>12</sup> In their work, they applied an automated classification algorithm to identify RCM image regions as either benign nevi lesions or melanocytic skin tumors. Their algorithm classified  $46.71 \pm 19.97\%$  of the

\*Contributed equally to this work, i.e., shared senior authorship.

Address all correspondence to: Sila Kurugol, Northeastern University, Electrical and Computer Engineering, 360 Huntington Avenue, Boston, MA 02115. Tel: 617-373-5191; Fax: 617-373-8970; E-mail: kurugol.s@neu.edu.



**Fig. 1** Left figure shows the DEJ in a vertical histology cross-section image and the middle and right figures show lateral slices from a RCM stack with the epidermis/dermis boundary marked. The DEJ is a thin membrane, shown with a blue solid line, that separates the epidermis from the dermis, a single layer of basal cells lies directly on the DEJ. The basal cell layer is typically at average depth of  $100\ \mu\text{m}$  below the surface in normal skin and  $10$  to  $15\ \mu\text{m}$  in thickness (Ref. 1). (Color online only.)

tumor images in benign melanocytic skin lesions as “malignant,” in contrast to  $55.68 \pm 14.58\%$  in malignant melanocytic skin lesions.

In this paper, as an initial target, we present a method for semiautomatic localization of the irregular three-dimensional dermis/epidermis junction (DEJ) in RCM images of human skin. The DEJ separates the superficial epidermis from the underlying deeper dermis. We chose to study the localization of the DEJ for three major reasons. First, the DEJ is a clinically meaningful target because cancers such as melanoma and basal cell carcinoma both originate and later spread from this junction. Thus, clinicians (when examining patients) and pathologists (when examining biopsy sections) need to accurately and repeatably evaluate the DEJ for screening, diagnosis, and staging of skin cancers and lesions. Secondly, the DEJ is also a clinically significant structure in many other types of epithelial tissue. Finally, we believe that localization of the DEJ is a useful surrogate for other desired forms of clinically relevant automation of RCM image analysis. Our findings may then be expected to inform and guide future attempts to detect and classify pathologies in skin and also in other tissues.

In dark skin or strongly pigmented skin types, melanin pigment in the basal cell layer creates a strong contrast. The basal cell layer lies directly on the DEJ. Thus, localization of the pigmented basal layer may offer a useful surrogate for the DEJ. By comparison, localization of the DEJ is difficult in RCM images of fair or lightly pigmented skin types due to lack of melanin pigmentation and lack of contrast at the basal layer. Indeed, this is a key motivation for this work: in this initial effort (of an anticipated long-term study), our choices were to address either strongly or lightly pigmented skin types. We chose to address the more difficult cases of lightly pigmented skin. This choice was carefully made based on advice from our clinical colleagues in dermatology.

This difficulty is compounded by high intra/intersubject, and even intralayer variability, resulting from the natural biological heterogeneity of skin tissue. In particular, the epidermis layer has a changing depth-dependent layered structure composed of cells with different morphology at each layer, while the dermis is mostly collagen fibers and blood vessels and is inconsistent in appearance. Due to heterogeneity of skin, at some sites dermis region including collagen fibers appear very bright and at other sites, dermis region appears dark, and at some sites dermis regions looks very much like epidermis. Moreover, since the DEJ is highly corrugated with a hills-and-valleys topography (Fig. 1), its location is not a simple function of depth from the skin sur-

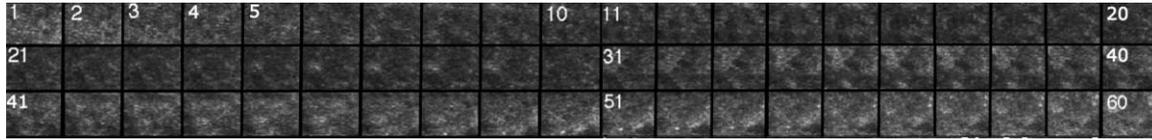
face. This topography of the DEJ surface also causes a lack of consistent appearance and a lack of discriminative features of interest for important local regions, especially regions where the DEJ surface has a large slope. Experience with translational and clinical studies with RCM has shown that clinicians often cannot visually find a single well-defined and consistent junction with accuracy and repeatability without a significant amount of training; even with training, experts may find it difficult to reliably locate the DEJ.

Indeed, after our steps in initial algorithm development<sup>13</sup> and subsequent feedback from clinical collaborators, we determined that localization of a single DEJ boundary (i.e., a single DEJ) is not consistently supported by the data. Instead, the goal adopted here is to locate two 3D surfaces that together bound a transition region in which the DEJ is located. The region above the first surface is determined to be epidermis with high confidence, and the region below the second surface is determined to be the dermis, also with high confidence. We reported some preliminary results in Kurugol et al.<sup>14</sup>

In this paper, we report further advances to solve this DEJ localization problem. To attack this problem, we found it necessary to incorporate ideas from three fields: texture segmentation, pattern classification, and sequence segmentation (SS). One component of our algorithm grew out of a class of standard approaches to segment regions based on their different textural features by using a classifier<sup>15</sup> (see Ref. 16 for a general review of texture segmentation algorithms). Beyond such standard 2D texture segmentation, we incorporated ideas from dynamic texture segmentation and texture change detection. Dynamic texture segmentation<sup>17</sup> is used to segment a sequence of 2D images with identical spatial statistics but dynamic changes along a third dimension. This dimension can be time, as in image sequence segmentation; here, the third dimension is depth through the skin, along the optical axis. Such problems can be solved by sequence segmentation algorithms<sup>18</sup>. For example, a change detection solution for image sequences is reported in Ref. 19.

Our algorithm works on small regions or “tiles.” The algorithm partitions each vertical stack of tiles (see Fig. 2 for an example) into a variable number of homogeneous segments in  $z$  (depth). We define homogeneity in terms of a multivariate dynamic model on a chosen set of image texture features. Based on the resulting set of break points between segments, we then identify two of them as the transition region boundaries using a multiclass classifier in the  $z$ -direction that assigns segments to epidermis, transition, or dermis. This process starts from the upper and lower images in each stack and moves toward the middle until the classifiers locate the respective region boundary at a segment end point.

We evaluated two different training scenarios for our algorithm. The first scenario was based on training and testing the algorithm on the same stack. In this scenario, we labeled a small training set for a given stack and used it to train the classifiers, and then we applied them to the rest of that stack (same stack scenario). In the second scenario, we labeled and trained on an entire stack and applied the resulting classifiers to different stacks from subjects with similar skin types (cross stack scenario). Results from both scenarios showed a reasonable performance considering the difficulty of the problem. Our methods and results are reported below.



**Fig. 2** An example stack (sequence) of 60 tiles is shown, with increasing depth indicated by increasing slice number in the figure. For this stack, an expert evaluator (see Sec. 3 for details) located the epidermis boundary at slice 19 and the dermis boundary at slice 29.

## 2 Methods

In this section, we describe our algorithm for the semiautomatic localization of the DEJ in RCM images. The stages of the algorithm are summarized in the flow chart (Fig. 3). We will describe each of these stages in turn. In particular, in Sec. 2.1, we describe the acquisition and preprocessing of the RCM data stack. Section 2.2 describes the set of texture features extracted from the training data and the feature selection algorithm used to select the most relevant, least redundant features. We explain sequence segmentation and classification algorithms employed in the two parallel stages in Secs. 2.3 and 2.4, respectively. In Sec. 2.5, we explain how the algorithm combined the results of the sequence segmentation and classification stages to identify which segment boundaries correspond to skin layer boundaries. Finally, we describe some post-processing applied to smooth the two resulting boundaries in Sec. 2.6.

### 2.1 Data Acquisition and Preprocessing

Acquisition of stacks of images of human skin *in vivo* was performed with a commercial reflectance confocal microscope (VivaScope 1500, Lucid Inc., Rochester, New York). The design and instrumentation details were reported earlier.<sup>20</sup> Briefly, the tissue was illuminated with a near-IR diode laser at a wavelength of 830 nm at low power (5 to 10 mW). For localization of DEJ, we are not restricted to stacks which start from the surface of the skin. It is sufficient for the algorithm that the stacks start from any depth within the epidermis layer and terminate at any depth in the dermis. Hence, each stack was acquired by starting from an arbitrary location in the epidermis, capturing the first slice, and then successively moving the focal plane 1- $\mu\text{m}$  deeper, until 60 slices were acquired. The laser power for each slice was automatically adjusted such that the pixel intensity range of a certain slice covered the complete 8-bit dynamic range of the imaging system. An RCM stack from a single skin site consisted of 60 image slices, where each slice was  $1000 \times 1000$  pixels, with a pixel resolution of 0.5  $\mu\text{m}$ . We note that although the  $z$  step resolution was 1  $\mu\text{m}$ , the true optical sectioning thickness of the imaging system was 3  $\mu\text{m}$ .

In each stack, we manually masked out gross undesired structures such as wrinkles. We then registered the slices in the transverse direction at different depths, again using a standard method, normalized cross-correlation.<sup>21</sup> For further processing, we divided each image in the stack into  $50 \times 50$  pixel ( $25 \mu\text{m} \times 25 \mu\text{m}$ ) tiles. From this point on, we performed tile-wise processing instead of pixel-wise processing, primarily for computational efficiency.

### 2.2 Feature Extraction and Automatic Supervised Feature Selection

We represented each image tile by a set of features that we hypothesized would be important for discriminating among epidermis, dermis, and transition regions. We extracted a large number (170) of such texture features from each tile, including gray level co-occurrence matrix features (contrast, energy, correlation, and homogeneity), statistical metrics (mean, variance, skewness, and kurtosis), features from a wavelet decomposition,<sup>15</sup> log-Gabor features, and radial spectral features. Table 1 gives a complete list of features. Similar features were found to be useful and explained in detail in Ref. 22. We obtained labeled training data for feature selection and classifier training by manual labeling of either a partial or a complete stack using one of the two different learning scenarios described in Sec. 1 (i.e., same stack and cross stack training scenarios).

We ran an automatic feature selection algorithm twice on the labeled training data: In both cases, the goal was to select a subset of features from our full feature set by choosing the most discriminative and least redundant features using the training set of labeled tiles from each class. In the first run, the two most discriminative and least redundant subset of features, one for each class (epidermis or dermis) against other classes were selected. These three subsets of features were used for training the classifiers. In the second run, the goal was to select features for the sequence segmentation algorithm: The most discriminative and least redundant subset of features for both classifiers was selected by finding the union of the two subsets of features in the first run.



**Fig. 3** Flow chart of the algorithm.

**Table 1** The complete set of features.

Category	Details	Number
Graininess feature	mean contrast difference of the original image and the wiener filtered (blurred) image	1
Features from cooccurrence matrix (Ref. 23)	contrast, correlation, energy features for diagonal and vertical neighbors	2×3
Statistical features (Ref. 15)	standard deviation, kurtosis, skewness	3
Wavelet packet decomposition features (Refs. 15 and 24)	energy and variance from 64 subimages (3-level decomposition using Daubechies-10 wavelet)	64×2
Log-gabor filter features (Ref. 25)	mean and variance over all orientations as a function of scale (4-scales)	4×2
Radial spectrum features (Ref. 21)	spectral energy as a function of radius	24

Specifically, for both runs, we applied a fast supervised feature selection algorithm based on a fast filter method.<sup>26</sup> The fast filter method searches for the best features one at a time. It has two steps: 1. Rank and select the features by how relevant they are for distinguishing different classes, based on Fisher's class separation distance measure (defined below), and 2. find the subset of relevant features that are least redundant to other relevant features based on a correlation measure.

The Fisher's class separation distance used here, for feature  $x$  between class 1 and class 2,  $D_x(c_1, c_2)$ , is given as follows:

$$D_x(c_1, c_2) = \frac{|\mu_{c_1} - \mu_{c_2}|}{\sqrt{\sigma_{c_1}^2 + \sigma_{c_2}^2}}, \quad (1)$$

where  $\mu_{c_1}$  and  $\mu_{c_2}$  are the mean values and  $\sigma_{c_1}$  and  $\sigma_{c_2}$  are the standard deviations of the features over class 1 and 2, respectively.

The fast filter method searches for the best features one at a time. The first step selects the features with normalized Fisher's distance (obtained by dividing Fisher's distance by the largest Fisher's distance in the feature set) larger than an experimentally chosen threshold of 0.25. The second step removes relatively correlated features from the selected set. Specifically, it first creates an ordered list of features by ranking the features in descending order according to their Fisher's distance. It starts from the first feature in the list, i.e. the feature with largest  $D_x$ , say  $F_i$ , and it removes features that have a correlation measure with  $F_i$  larger than another experimentally chosen threshold of 0.9. Then, it sets  $F_i$  to the next remaining feature in the list and repeats the procedure until all remaining features in the list are visited.

### 2.3 Sequence Segmentation in $z$ -Direction

We now explain the two parallel stages used to extract boundaries and classifiers, which are subsequently combined to determine boundaries. We describe the sequence segmentation algorithm, which uses the features selected for that algorithm as described above. To start, we denote that set of  $d$  features for all  $n$  tiles in a given stack (or sequence) of tiles as  $F = \{f_1, \dots, f_n\}$ , of length  $n$ , where each point is  $d$ -dimensional. A segmentation

partitions  $F$  into  $k$  contiguous segments such that each segment is as homogeneous as possible as in the sense below.

We applied the sequence segmentation algorithm to each sequence of tiles in the 3D RCM stack.

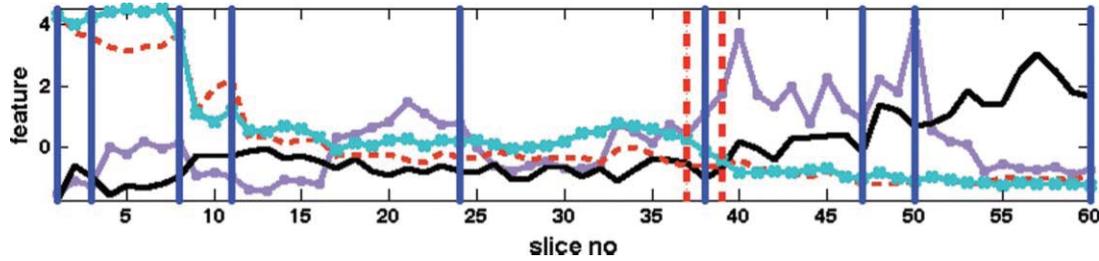
In detail, for each  $k$  in a fixed range (here,  $k$  takes values 3, 4, ..., 8), using an optimal dynamic programming (DP) algorithm, we partitioned the multidimensional feature vector sequence into  $k$  homogeneous segments. We modeled each dimension of our  $d$ -dimensional feature vector with a piecewise affine model for each segment. The cost function used by the DP algorithm is the  $l_2$  norm of the error between that model and the features:

$$\text{Cost} = \sum_{j=1}^k \sum_{i \in s_j} \|\mathbf{f}_i - (\mathbf{a}_j i + \mathbf{b}_j)\|_2. \quad (2)$$

Here, given the desired number of segments  $k$ , the depth indices of the tiles, denoted by  $i$ , are partitioned into a set of  $k$  contiguous subsets, denoted  $s_j$ , one for each segment. For each  $i$ 'th tile within a segment, the  $d$ -dimensional feature  $\mathbf{f}_i$  ( $\mathbf{f}_i = [f_{1i}, f_{2i}, \dots, f_{di}]$ ) is approximated by the affine combination  $\mathbf{a}_j i + \mathbf{b}_j$ , where the  $d$ -dimensional vectors of affine function parameters  $\mathbf{a}_j$  ( $\mathbf{a}_j = [a_{1j}, a_{2j}, \dots, a_{dj}]$ ) and  $\mathbf{b}_j$  ( $\mathbf{b}_j = [b_{1j}, b_{2j}, \dots, b_{dj}]$ ) are held constant throughout the  $j$ 'th segment but the affine "variable"  $i$  varies with tile index. We collect the approximation errors across the  $k$  segments as our cost. We then optimized for the model parameters for each segment along with the segment boundaries (definitions of  $s_j$ ), for a given  $k$ , by jointly minimizing the cost across all variables.

We used a heuristic approach to find the number of segments to use: We first ran the algorithm for  $k$  from 3 to 8. Then the cost ratio between using  $k - 1$  and  $k$  segments was calculated for  $k$ 's starting from 4 and repeating until this ratio fell below a threshold  $t$  (set to 1.3 experimentally).

In Fig. 4, we show an example multivariate  $z$ -sequence of features. For illustration purposes, only four features were plotted. Each feature is represented by a different color. The  $z$ -sequence is of length 60, i.e., 60 slices. The segment boundaries of the eight segments found by the sequence segmentation algorithm are shown with solid blue vertical lines. The dashed vertical red



**Fig. 4** An example multivariate z-sequence of features. For illustration purposes, only four features are shown. The segment boundaries of the eight segments found by the sequence segmentation algorithm are shown with solid blue vertical lines. The dashed vertical red lines show the epidermis and dermis boundaries located by the expert. (Color online only.)

lines show the epidermis and dermis boundaries located by the expert. We note that one can visually observe a qualitative difference between each adjacent pair of regions, which presumably is what determined where the sequence segmentation algorithm placed the boundaries automatically.

#### 2.4 Classifier Training (Locally Smooth Support Vector Machine)

We used a modification of a standard two-class classification technique known as a support vector machine (SVM). In a SVM classifier design, each point, represented by  $d$  features, is treated as a point in  $R^d$ . A hyperplane is used to separate data points in one class from those in the other. The hyperplane is chosen so as to have the largest possible distance to the nearest training data points of any class, since, in general, that hyperplane will lead to better generalization of the classifier to new examples.<sup>27</sup> Here, we measure the 1-norm distance between the hyperplane and the data points.

A standard SVM classifier assumes all samples are independent of each other. Since the class (layer) of adjacent tiles in the transverse plane will clearly be significantly correlated, we found it useful to leverage that correlation in our classifier. Therefore we employed a variation of SVM known as locally smooth SVM (LS-SVM). The LS-SVM algorithm<sup>28</sup> is a recently introduced modification of a standard SVM which takes into account the spatial correlation of samples by allowing correlated points to affect the classification of the current point. To do so during training an additional term is added to the standard SVM optimization function. This additional term changes the classifier such that neighboring tiles are more likely to be classified as belonging to the same class. During classification, a tile was provisionally classified using the trained preclassifier and then a final decision was obtained by also considering the weighted decisions of neighboring tiles. Our implementation is similar to that in Ref. 28 except that for the matrix  $R$  that represents the additional term, we used a block diagonal matrix having a  $4 \times 4$  matrix of ones as its block. This matrix enforces spatial correlation by equally weighting a (casual quarter-plane) neighborhood of three adjacent tiles.

We trained two different 1-norm LS-SVM classifiers: One for epidermis versus nonepidermis (epi-rest) tiles and one for dermis versus nondermis (der-rest) classes using the features selected with the method explained earlier. We took the distance from each point (here, the feature vector of a particular tile) to the classifier hyperplane as a decision metric for that tile

(where negative numbers were used on the “wrong” side of the hyperplane). To enable comparison between classifiers, we normalized this distance measure for each classifier so that it ran from 0 to 1 by passing it through a logistic function. For example, a distance of 0.5 meant that both classes were equally likely. We can consider this normalized distance as a probability measure of belonging to a class. We are now in a position to combine this measure with the results of the dynamic sequence segmentation stage.

#### 2.5 Combined Sequential Decision

The basic approach we adopted to combine these two sets of results is to use the sequence segmentation endpoints as the candidate set of final layer boundaries for each tile. We used the classifiers to determine which two endpoints to choose as those boundary locations. In particular, we employed the following procedure once from top to bottom for the epi-trans boundary and again from bottom to top for the trans-der boundary.

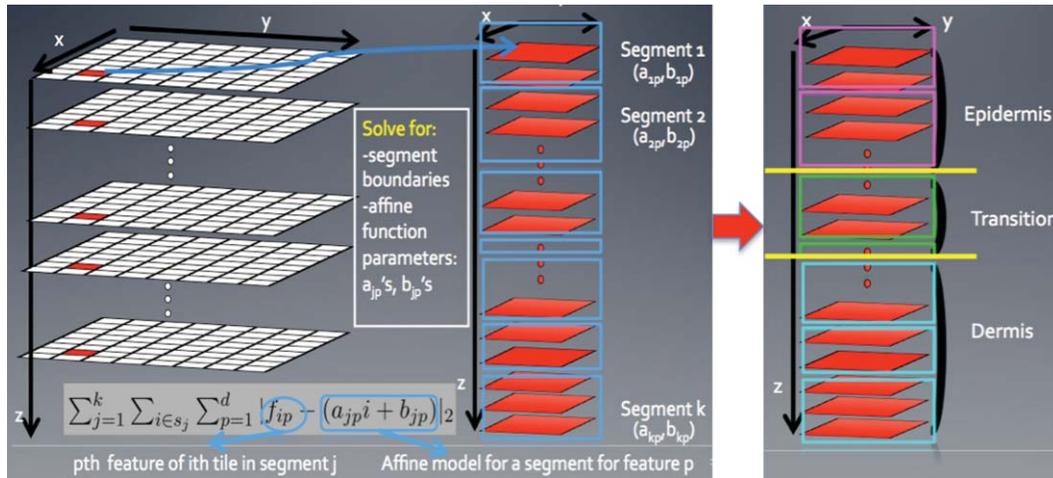
Starting from the top (bottom) segment of tiles in the stack we identified the epi-trans (trans-der) boundary as the segment end point beyond which the average probability of belonging to epidermis (dermis) from the epidermis versus rest (dermis versus rest) classifier was below 0.4. This number was selected to be less than 0.5 intentionally to account for the noise and the effect of averaging the probabilities, i.e., if probabilities of most of the tiles within a segment are around 0.5 but a few tiles in that segment are below, we intend to include that segment to that class. Figure 5 illustrates the procedures of sequence segmentation and boundary location decision. For the sequence in Fig. 4, the algorithm located the epidermis boundary and dermis boundary both at the second segment boundary (with slice number 37).

#### 2.6 Smoothing and Post-Processing

Finally, we applied a smoothing filter to each of the two boundary surfaces located by the algorithm. Specifically, we used a Gaussian smoothing filter of size  $5 \times 5$  with standard deviation 0.75 to each boundary, where these dimensions given are in units of tiles (note that each image slice is comprised of  $20 \times 20$  tiles).

### 3 Experiments

As explained in Sec. 1, we used two different scenarios for training. In the same-stack-training scenario, out of 41 tile sequences



**Fig. 5** Left panel shows the tile sequence and an example output of the sequential segmentation algorithm. Right panel shows the resulting epidermis and dermis boundaries (yellow longer horizontal lines) of the combined sequential+classification decision algorithm. (Color online only.)

in the stack, 8 tile sequences were used, 5 for validation and 3 for training, to train our classifier and tune the classifier parameters. We performed standard  $N$ -fold ( $N = 6$ ) cross validation experiments to tune our classifier parameters. In these experiments, a “brute force” search in the parameter space was performed, followed by choosing the parameter combination with the best classification performance over the validation set. In order to provide the LS-SVM algorithm with the neighboring samples it uses for local smoothing, tiles were chosen in  $2 \times 2$  adjacent blocks, of size  $100 \times 100$  pixels, which were then subdivided into  $50 \times 50$  pixel tiles. The reason for choosing this small neighborhood size of  $2 \times 2$  is due to a constraint of the expert markup in scenario 1. In particular, in scenario 1 the expert only marked a small number of training samples. To ensure the availability of adjacent tiles for training, we had the experts mark regions of size  $100 \times 100$  pixels (giving us  $2 \times 2$  neighborhoods of  $50 \times 50$  tiles). The expert marked the two boundary depths in these selected tile stacks. The rest of the processing was automated once these labeled tiles were provided. In scenario 2, the cross-stack-training scenario, we assumed that data stacks from similar skin types have similar features. Thus, a classifier pretrained on one completely labeled RCM stack from a particular skin type was applied to a new RCM stack from a similar skin type without any need to label tiles in the new stack.

#### 4 Results

We report results from experiments on four RCM test data sets from four different subjects with fair skin types (Ref. 29)—skin type I and II—which as noted above are the most difficult cases due to the small amount of melanin pigment and, hence, contrast. Results are reported for both training scenarios (scenario 1: same stack and scenario 2: cross stack). Note that the expert fully labeled both stacks. In both scenarios, the expert markings not made available to the algorithm were used to evaluate the performance.

Since in scenario 1 the expert only marked a small number of training samples, we made sure that marked regions included at least one neighborhood; to do so we had the experts mark regions of size  $100 \times 100$  pixels (giving us  $2 \times 2$  neighboring

$50 \times 50$  tiles). Thus, we used a small neighborhood of size four (three adjacent tiles). Note that scenario 2 does not require this restriction, and as the reviewer pointed out, it would make more sense to use 4 or 8. To be consistent for both training scenarios, we utilized the neighborhood of size four (three adjacent tiles) consistently in the results reported here.

**Table 2** Results of Scenario 1 and Scenario 2 are given for both RCM stack 1 (column 2) and RCM stack 2 (column 3). Rows labeled N give the number (and ratio) of test tiles (out of the total tested) for which the detected boundary was within  $10 \mu\text{m}$  (N10) or  $15 \mu\text{m}$  (N15) of the expert marked boundary. Rows labeled  $m \pm \text{std}$  give the mean and standard deviation of the error in micrometers between detected and expert marked boundaries across the test set.

	RCM stack 1	RCM stack 2
Scenario 1		
N10 (Epi.)	110/164 = 0.67	94/164 = 0.57
N10 (Der.)	117/164 = 0.71	101/164 = 0.62
N15 (Epi.)	137/164 = 0.84	136/164 = 0.83
N15 (Der.)	138/164 = 0.84	136/164 = 0.83
$m \pm \text{std}$ (Epi.)	$9.33 \pm 7.55$	$9.44 \pm 6.15$
$m \pm \text{std}$ (Der.)	$8.50 \pm 8.09$	$9.16 \pm 6.58$
Scenario 2		
N10 (Epi.)	125/164 = 0.76	111/164 = 0.68
N10 (Der.)	142/164 = 0.87	116/164 = 0.71
N15 (Epi.)	144/164 = 0.88	135/164 = 0.82
N15 (Der.)	150/164 = 0.92	138/164 = 0.84
$m \pm \text{std}$ (Epi.)	$7.94 \pm 5.93$	$9.51 \pm 7.13$
$m \pm \text{std}$ (Der.)	$6.72 \pm 4.82$	$8.75 \pm 8.33$

**Table 3** Scenario 1: Confusion matrices for RCM stack 1 RCM stack 2 as test set. Confusion matrices show the results of algorithm (columns) given the expert results (rows), as explained in the text.

		Algorithm		
		Epi.	Trans.	Der.
		RCM stack 1 as test set		
Expert	Epidermis	0.76	0.14	0.10
	Transition	0.47	0.33	0.19
	Dermis	0.10	0.10	0.80
		RCM stack 2 as test set		
Expert	Epidermis	0.64	0.27	0.09
	Transition	0.28	0.35	0.37
	Dermis	0.03	0.16	0.83

The results from the automatic supervised feature selection algorithm indicate that log-Gabor and wavelet features were the most commonly selected features. These features are consistent with the visual features used by an expert to discriminate epidermis from dermis, which are typically the blurriness of dermis versus the cellular texture pattern of epidermis. We speculate that this information can be captured well by obtaining frequency information localized in space provided by the features calculated from log-Gabor filtered and wavelet packet decomposed tiles. The energy and variance features calculated from the wavelet transform at different scales provided texture information at those scales. Log-Gabor filters, which are a product of a Gaussian (with frequency represented in log-scale) and a sinusoid, are also used for texture discrimination at various scales. For example, for structured cellular epidermis regions, we obtained higher values from our energy feature at high frequency bands, while for blurry collagen fibers within dermis, we obtained higher values from our energy feature at lower frequency bands. This kind of texture exploration at multiple scales provided us the information needed to discriminate epidermis from dermis. However given the variability of the data and the stochastic nature of the feature selection process, we cannot make definitive statements either about which features are selected (or even the number of features selected) nor a physical interpretation of this outcome.

To quantify performance we calculated the distance in micrometers between the expert labeled boundaries and classification boundaries on each test set. For each experiment, we calculated the number of  $z$ -sequences of tiles for which this distance was smaller than both 10 and 15  $\mu\text{m}$ . We report this number as well as the mean and standard deviation of the error distances for both stacks in Table 2 for scenarios 1 and 2, respectively. We also report the results as confusion matrices. The confusion matrices for the first and second scenarios are shown in Tables 3 and 4, respectively. The  $(j, k)$ 'th entry of the confusion matrix indicates the number of tiles found to be class  $k$  [1-epidermis, 2-transition (transition) or 3-dermis] by the algorithm

**Table 4** Scenario 2: Confusion matrices for RCM stack 1 RCM stack 2 as test set. Confusion matrices show the results of algorithm (columns) given the expert results (rows), as explained in the text.

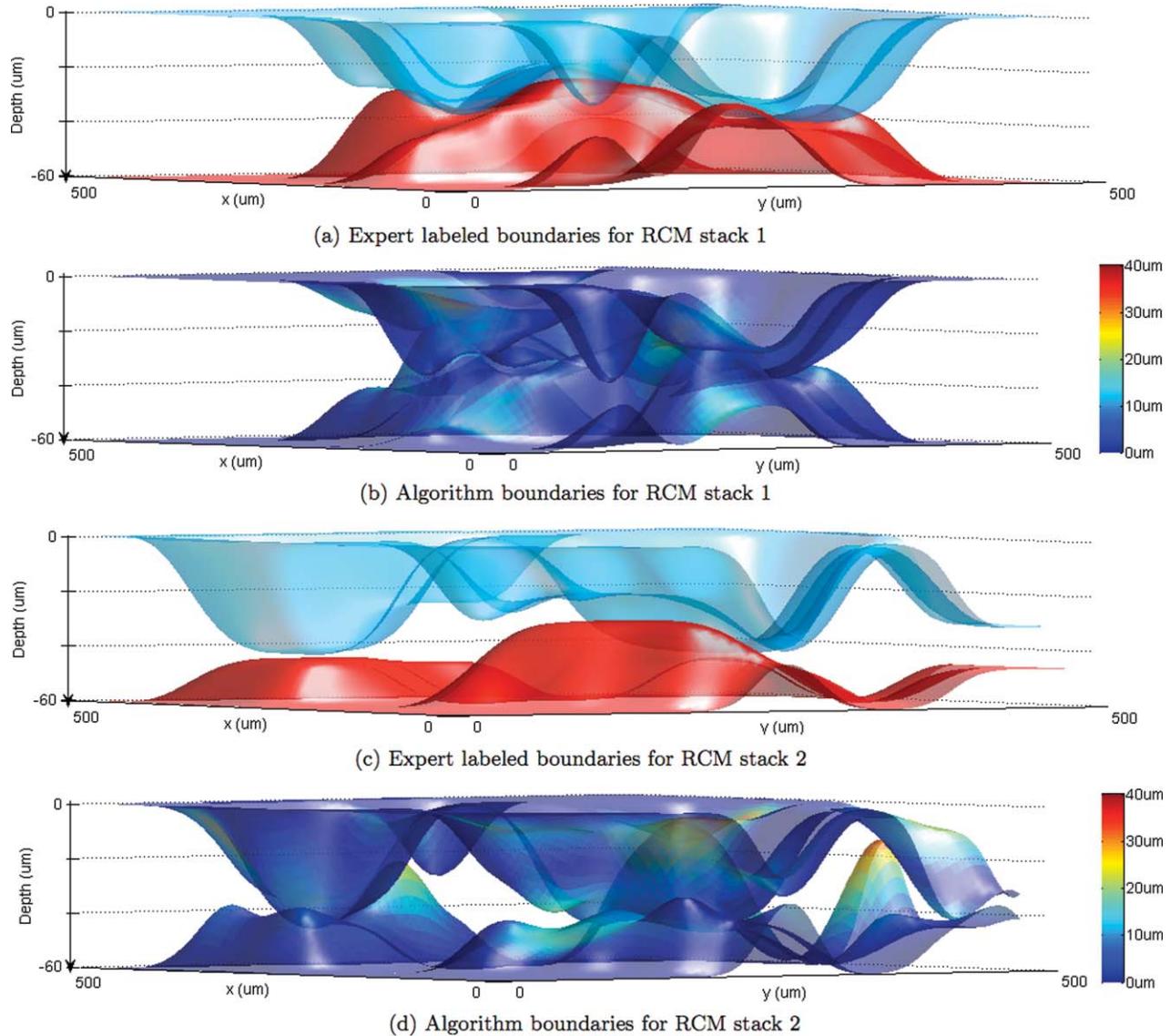
		Algorithm		
		Epi.	Trans.	Der.
		RCM stack 1 as test set		
Expert	Epidermis	0.66	0.28	0.06
	Transition	0.11	0.52	0.37
	Dermis	0.02	0.21	0.77
		RCM stack 2 as test set		
Expert	Epidermis	0.63	0.30	0.07
	Transition	0.14	0.57	0.29
	Dermis	0.03	0.15	0.76

given that they belong to class  $j$  according to expert markings. For both data sets we also plot the resultant 3D epidermis and dermis boundary surfaces for scenario 2 in comparison to the expert marked boundary surfaces in Fig. 6. Single-color surfaces [Figs. 6(a) and 6(c)] show the expert labeled epidermis (dermis) boundary while the multicolored surfaces [Figs. 6(b) and 6(d)] indicate the boundaries found by the algorithm. The color maps on the multicolored surfaces show the distance from the expert labeled boundary. In Fig. 7, for both RCM stacks and scenarios 1 and 2, we show the epidermis and dermis boundaries located by the algorithm in comparison to the expert located boundaries for all 164<sup>†</sup>  $z$ -sequence of tiles. In Fig. 8, a comparison of expert markings with the smoothed algorithm results are shown for scenario 1 in two orthogonal vertical views at the locations from the RCM stack 1 indicated by the solid lines drawn in the axial views (on the left).

We also show average results for experiments performed on stacks 1 to 4. In those experiments, a classifier is trained on one stack and applied to the remaining three stacks. Therefore, three different classifiers were applied to each stack and the results were calculated. The results for each stack averaged over all three classifiers applied to that stack are reported in Tables 5 and 6. Table 5 shows average confusion matrices for RCM stacks 1 to 4. The average was calculated over three classifiers applied on a stack. For each stack, the classifier trained on that same stack was not used in the testing. Table 6 also reports average results calculated similarly. Rows labeled  $N$  show the mean ratio of test tiles (out of the total tested) for which the detected boundary was within 10  $\mu\text{m}$  (N10) or 15  $\mu\text{m}$  (N15) of the expert marked boundary. Rows labeled  $m \pm \text{std}$  give the average mean and standard deviation of the error in micrometers between detected and expert marked boundaries across the test set.

The results from the same experiments are also reported by averaging the performance of each classifier over the three

<sup>†</sup>Initially 400 tiles of size  $50 \times 50$  pixels were obtained by dividing a slice of  $1000 \times 1000$  pixels into tiles. We only processed 164 tiles out of these 400 tiles, since tiles which included wrinkles were masked out during preprocessing.



**Fig. 6** Surface plot of the epidermis boundary and the dermis boundary in 3D in comparison to the expert labeled boundaries of RCM stack 1 and 2 for scenario 2 (cross training). Top blue (bottom red) surfaces show the expert labeled epidermis (dermis) boundary for (a) RCM stack 1 and (c) RCM stack 2. The colored surfaces indicate the resultant boundaries of the algorithm for (b) RCM stack 1 and (d) RCM stack 2. The color maps indicate the distance from the expert labeled boundary. The z-axis is in micrometers. x and y axes are in pixels, where the pixel spacing is  $0.5\mu\text{m}$ . Flat regions are the masked out wrinkles. For the smooth visualization purpose, the boundaries are plotted after interpolating them twice in 2D with spline interpolation. (Color online only.)

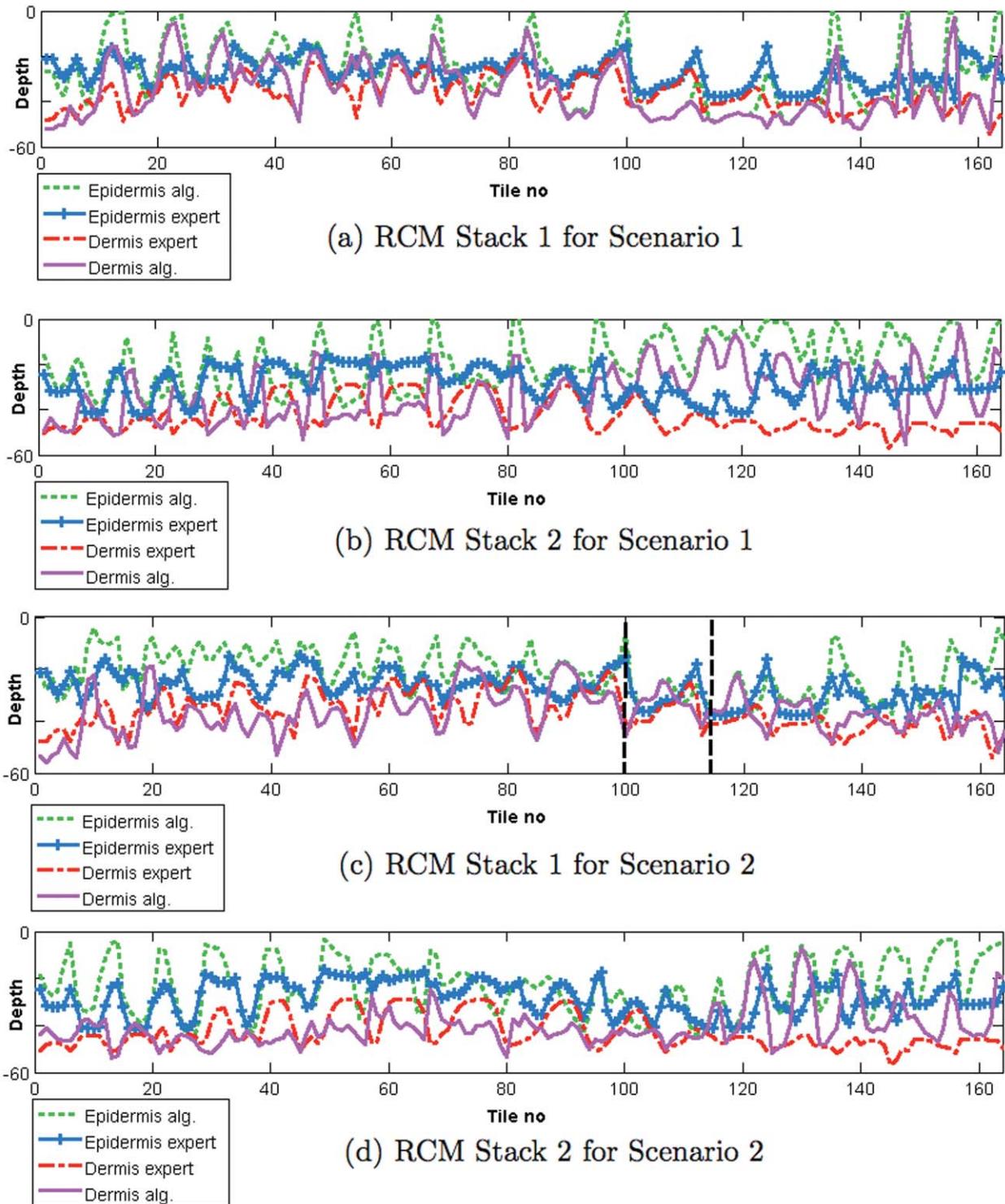
stacks that it was tested on. For each classifier, the stack on which the classifier was trained on was not used in the testing. Table 7 shows average confusion matrices for classifiers 1 to 4 and Table 5 shows average distance results calculated similarly for each classifier. Thus with these tables we can examine both the performance of each classifier across all other stacks and the performance of all classifiers on each stack.

The video, which shows the classification results of scenario 2 applied on RCM data stack 1, is provided as a multimedia file and a snapshot from the video is shown in Fig. 9. The left panel shows epidermis and dermis surfaces and a cutting data slice that moves from the top of the stack to the bottom. The right panel shows the original data slice (bottom) and the same slice with the overlaid algorithm results (top). The video starts from a superior slice of the stack, where all regions were either

classified as epidermis (red shaded) or were masked out (dark gray shaded) in the preprocessing stage. Then the cutting plane proceeds to deeper slices. Moving deeper in the stack, first the epidermis regions shrink and the transition regions (light gray shaded) start. Then the transition regions shrink and the dermis regions (blue shaded with solid boundary) start. The deepest slices in the stack include only dermis regions.

## 5 Discussion

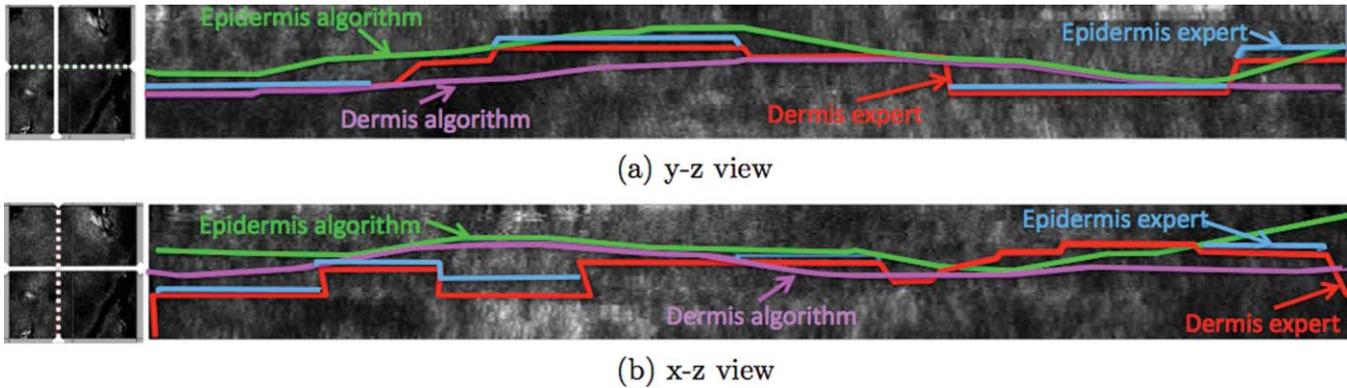
In this work, we developed an algorithm to locate the DEJ in RCM image stacks for lightly pigmented skin types. The proposed hybrid algorithm locates the epidermis and dermis boundary surfaces with a decision combining sequential segmentation and classification stages. The results show that the algorithm



**Fig. 7** For scenario 1 and 2 and RCM stacks 1 and 2, the figure show the epidermis and dermis boundaries located by the algorithm in comparison to the expert located boundaries for all of the 164 tile-sequences that were processed by the algorithm. The boundaries shown are 2D Gaussian filtered for smoothness as explained in the post-processing step in Sec 2. The dotted vertical lines in (c) indicate the location of the vertical slice shown in Fig. 8(b). (Color online only.)

performed reasonably well with epidermis/dermis misclassification rates smaller than 10% and average distance from the expert labeled boundaries around 7 to 12  $\mu\text{m}$ . When considering the fact that the standard epidermis—dermis classification

approaches perform poorly due to a number of reasons including the heterogeneity of skin layers, the proposed combined decision algorithm achieves a significant improvement by making use of the spatial structure and dynamics inherent in the tissue.



**Fig. 8** Comparison of expert markings with the algorithm results shown in vertical views  $y$ - $z$  (top) and  $x$ - $z$  (bottom). The solid line on the left of both (a) and (b) indicate the vertical slice location. Transition region is located by the algorithm in between epidermis algorithm (green) and dermis algorithm (purple) curves. The green (purple) curve is the epidermis (dermis) boundary found by the algorithm. The blue (red) curve is the dermis (epidermis) boundary marked by the expert. If there is no epidermis expert (blue), the expert found no transition region and the upper and lower boundaries coincide. For visualization purposes, algorithm boundaries computed for each tile are linearly interpolated to the same grid (pixel grid) that the expert used in their mark-up. (Color online only.)

Due to this heterogeneity of skin structure, for less than 10% of all tile stacks<sup>†</sup>, the classifier effectively could not distinguish between the different layers. Tiles in these tile-stacks were found upon visual inspection to have an ambiguous appearance which was dissimilar to the large majority of the tile stacks. Hence, the decision function that uses the LS-SVM classifiers failed to discriminate either epidermis from the rest or dermis from the rest due to lack of discriminative features between layers for these tile stacks and the concomitant lack of dynamics in the  $z$ -direction.

The results from all tile stacks are shown in comparison to the ground truth as determined by expert markups in Table 2. The presentation of our results implies that we accept an accuracy of within 10 to 15  $\mu\text{m}$  as useful. The rationale for using these thresholds is that because the size of a basal layer cell is about 10 to 15  $\mu\text{m}$ , we are within about one cell distance of the boundary. We note that when clinicians visually evaluate a data stack, they typically evaluate with 5  $\mu\text{m}$  slice separation, which is considered a reasonable distance for a visually detectable change to occur from slice to slice.<sup>30</sup> In that case, 10 to 15  $\mu\text{m}$  would correspond to 2 to 3 slices.

For the dermis boundary, as shown in Table 2, this acceptable distance from the expert boundary is achieved for more than 82% of all tile stacks. Similar results can be observed from the plots in Fig. 7 (i.e., the algorithm's dermis boundaries are almost touching the expert dermis boundaries for most of the tile stacks). The algorithm performance is reported in terms of the confusion matrices in Tables 3 and 4. Regions labeled as dermis both by the algorithm and the expert are given in the (3,3) entries of the confusion matrices, and again these values are around 80%. The epidermis/dermis misclassification rates for dermis are shown in (3,1) entries of the confusion matrices and are less than 10%. For experiments performed on all four RCM stacks, similar results averaged over a stack and averaged over a classifier were also reported in Tables 5–8. The average classifier performance on each stack reported in terms of confu-

sion matrices indicate a comparable performance over all stacks even though some classifiers might work better on some stacks and worse on other stacks due to variability from subject to subject. Mean distances between expert and algorithm boundaries were around 7  $\mu\text{m}$  for dermis and 9  $\mu\text{m}$  for epidermis and the acceptable distance of 15  $\mu\text{m}$  was achieved for around 90% of tiles for dermis and 85% of tiles for epidermis.

On the other hand, the epidermis boundaries of the algorithm are generally chosen more conservatively than the expert labeled epidermis boundaries. The algorithm was designed to label the regions as epidermis only if it finds high probability that they are epidermis; otherwise it labels the regions as transition. Hence, the epidermis boundary of the algorithm is on average about 20  $\mu\text{m}$  away from the expert labeled epidermis boundary. Consistent epidermis results are shown in Tables 3 and 4, where confusion matrix entries (1,1) indicate epidermis classification accuracies. These accuracies for epidermis are around 45%, while the (1,2) entries of around 40% indicate conservative detection of epidermis boundaries as regions labeled as epidermis by the expert are labeled as transition by the algorithm. Although the epidermis classification accuracy is not as high as for dermis due to this phenomenon of epidermis being classified as transition in many cases, the epidermis/dermis misclassification for epidermis, as shown in the (1,3) entries of the matrices, is very low, with values less than 10%, which suggests successful epidermis/dermis classification. If we decide, based on this analysis, to include the transition regions into the epidermis class—that is place our detected epidermis boundary below the transition region instead of above it—then the epidermis classification accuracy of the algorithm is increased to around 90% in all cases and dermis accuracy stays around 76%. In that case, the detected DEJ boundary would be a single surface, the dermis boundary found by the algorithm. Similar analysis for experiments performed on all four RCM stacks show that including transition regions into the epidermis results in accuracy values larger than 75% for dermis for all stacks and 90% for epidermis for all stacks except stack 3.

The transition region labeled by the algorithm generally includes some deeper epidermis regions lacking strong epidermis features and some superior dermis regions lacking strong

<sup>†</sup>Specifically we refer to tile stacks for which the sum of the distances from the expert label epidermis and dermis boundaries to those located by the algorithm was greater than 40 $\mu\text{m}$ .

**Table 5** Average confusion matrices for RCM stack 1 to 4 as test set. For each stack, the average was calculated over three classifiers trained on the other stacks and applied on the remaining stack. For each stack, the classifier trained on that same stack was not used in the testing. Confusion matrices show the results of algorithm (columns) given the expert results (rows), as explained in the text.

Test set	Algorithm											
	stack 1			stack 2			stack 3			stack 4		
	E	T	D	E	T	D	E	T	D	E	T	D
Expert Epidermis (E)	0.64	0.29	0.07	0.51	0.40	0.09	0.57	0.22	0.21	0.71	0.24	0.05
Transition (T)	0.24	0.54	0.22	0.07	0.56	0.37	0.27	0.34	0.39	0.20	0.59	0.21
Dermis (D)	0.05	0.21	0.74	0.01	0.19	0.80	0.05	0.12	0.83	0.03	0.28	0.69

**Table 6** Average results for stack 1 to stack 4 as test set. For each stack, the average was calculated over three classifiers trained on the other stacks and applied on the remaining stack. For each stack, the classifier trained on that same stack was not used in the testing. Rows labeled N show the mean ratio of test tiles (out of the total tested) for which the detected boundary was within 10 μm (N10) or 15 μm (N15) of the expert marked boundary. Rows labeled m ± std give the average mean and standard deviation of the error in μm between detected and expert marked boundaries across the test set.

Test set	stack 1	stack 2	stack 3	stack 4
N10 (Epi.)	0.62	0.47	0.70	0.83
N10 (Der.)	0.76	0.63	0.81	0.83
N15 (Epi.)	0.79	0.68	0.89	0.99
N15 (Der.)	0.87	0.81	0.94	0.97
m ± std (Epi.)	10.04 ± 6.55	12.32 ± 7.55	8.31 ± 5.51	7.43 ± 3.76
m ± std (Der.)	7.70 ± 5.80	9.56 ± 7.64	6.48 ± 4.80	6.77 ± 4.34

dermis features. For such regions, the expert labeled transition region is likely somewhat subjective, i.e., the height of the transition region in z may change from expert to expert. However, experts generally try to mark the transition region as narrowly as possible. In particular, we speculate that experts may, in ef-

fect, add in more complex perceptual features to their decision process, or use information from a large number of neighboring tiles to make decisions in uncertain regions, even if the sufficient low-level textural clues are not present. This suggests that the algorithm could be refined in the future to, similarly, identify regions of larger uncertainty, and in those cases employ more complex features, rely more on spatial context, and/or use hierarchical classification approaches.

In the current version of our algorithm, the LS-SVM classifier leverages the spatial correlations between tiles. However, during the sequential segmentation of each tile stack based on dynamics, the spatial correlation between neighboring tile stacks is not used in the segmentation. Moreover, in the combined decision step, where we locate the epidermis and dermis boundaries as one of the segment end points, we again do not consider the variation of tile stack boundaries across stacks within a neighborhood; we only enforce smoothness in our post-processing step. These limitations will also be addressed in future development of our approach.

One significant limitation of the algorithm is parameter tuning and sensitivity. The SVM parameters are tuned automatically with a grid search over a labeled validation set. This is the method commonly used in the SVM literature.<sup>31</sup> There are also two thresholds in the feature selection algorithm that are also tuned by using the validation set. As future work, a more robust feature selection algorithm such as the one in Ref. 32 may replace the current fast filter approach.

**Table 7** Average confusion matrices for classifiers 1 to 4. The average was calculated over three out of four data stacks, on which the same classifier was applied. For each classifier, the stack on which the classifier was trained on was not used in the testing. Confusion matrices show the results of algorithm (columns) given the expert results (rows), as explained in the text.

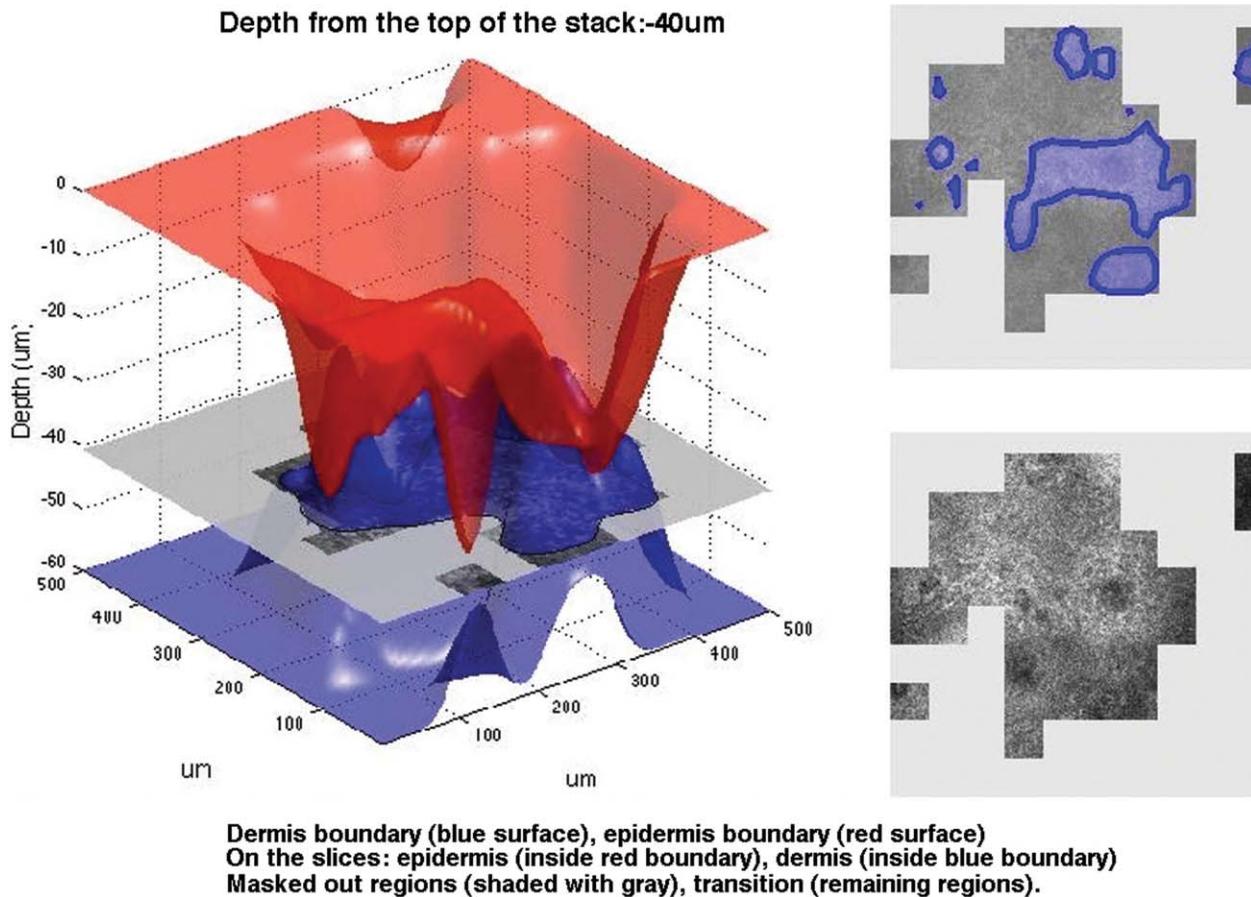
Classifier	Algorithm											
	stack 1			stack 2			stack 3			stack 4		
	E	T	D	E	T	D	E	T	D	E	T	D
Expert Epidermis (E)	0.67	0.23	0.10	0.67	0.22	0.11	0.64	0.28	0.08	0.46	0.41	0.12
Transition (T)	0.22	0.46	0.32	0.26	0.47	0.27	0.26	0.44	0.30	0.03	0.67	0.30
Dermis (D)	0.03	0.20	0.77	0.04	0.20	0.76	0.06	0.19	0.75	0.01	0.22	0.77

**Table 8** Average results for classifier 1 to classifier 4. The average was calculated over three out of four data stacks, on which the same classifier was applied. For each classifier, the stack on which the classifier was trained on was not used in the testing. Rows labeled N give the mean ratio of test tiles (out of the total tested) for which the detected boundary was within 10  $\mu\text{m}$  (N10) or 15  $\mu\text{m}$  (N15) of the expert marked boundary. Rows labeled m  $\pm$  std give the average mean and standard deviation of the error in micrometers between detected and expert marked boundaries across the test set.

Classifier	stack 1	stack 2	stack 3	stack 4
N10 (Epi.)	0.78	0.79	0.57	0.47
N10 (Der.)	0.80	0.82	0.68	0.72
N15 (Epi.)	0.92	0.91	0.82	0.70
N15 (Der.)	0.93	0.94	0.86	0.86
m $\pm$ std (Epi.)	7.62 $\pm$ 5.29	7.55 $\pm$ 5.08	10.45 $\pm$ 5.88	12.45 $\pm$ 7.10
m $\pm$ std (Der.)	6.95 $\pm$ 5.51	6.63 $\pm$ 4.76	8.79 $\pm$ 6.34	8.13 $\pm$ 5.96

We have begun to test our algorithm on a larger database, since the proof of concept reported in this paper indicated that training on one stack and testing on another stack from a similar skin type (cross training scenario) will likely be a useful method. This will be especially important if we wish to process a large number of RCM stacks without the need to mark up a training set in each individual stack.

In this work, we only treated data stacks from subjects with fair skin types, which have only a small amount of melanin pigment and hence almost no contrast at the DEJ location. These were, by far, the most challenging cases and we felt it was important to establish whether this approach could work under these low contrast conditions. For darker skin types, the availability of strong contrast at the basal layer provides a clearly detectable surrogate feature to localize the DEJ. In such skin, an algorithm based on peak detection in  $z$ -profiles of mean intensities of each tile stack may be adequate. A peak in the intensity profile is likely to be localized in the basal layer (located right above DEJ). In preliminary studies on a few cases it appeared that the first strong jump in that  $z$ -profile corresponded to the superficial stratum corneum layer, provided that the stack started from the skin surface, and the second strong jump corresponded to the



**Fig. 9** A snapshot from the video file which shows the classification results of scenario 2 applied on RCM data stack 1. The left panel shows epidermis and dermis boundary surfaces and a cutting data slice that moves from the top of the stack to the bottom. The right panel shows the original data slice (bottom) and the same slice with the overlaid algorithm results (top). The video starts from a superior slice of the stack, where all regions were either classified as epidermis (red shaded) or were masked out (dark gray shaded) in the preprocessing stage. Then the cutting plane proceeds to deeper slices. Moving deeper in the stack, first the epidermis regions shrink, and the transition regions (light gray shaded) start. Then the transition regions shrink and the dermis regions (blue shaded) start. The deepest slices in the stack include only dermis regions. (MPEG, 21.1MB) [URL: <http://dx.doi.org/10.1117/1.3549740.1>]

basal layer location superficial to the DEJ. We note that a similar idea was used in Ref. 33 to detect pagetoid melanocytes.

In our future work, we plan to test the DEJ localization algorithm in skin lesions and cancers *in vivo*. After locating the DEJ, the location of the lesion can be identified with respect to DEJ location. A similar or extended set of texture features can be calculated from the lesions and physical and physiological interpretation of most relevant features can be examined.

### Acknowledgments

The authors thank Dr. Allan Halpern and Dr. Alon Scope for clinical guidance, Dr. Juliana Casagrande and Dr. Itay T. Klaz for the acquisition of RCM data stacks, and Volkan Vural for providing his implementation of the LS-SVM classification algorithm. Support for the work of S.K. and D.H.B. was provided in part by the NIH/NCCR Center for Integrative Biomedical Computing (CIBC), P41-RR12553-09. J.D. was also partly supported by NSF IIS-0347532. Support for M.R. was provided by NIH grant no. R01EB006947.

- H. M. Gloster and D. G. Brodland, "The epidemiology of skin cancer," *Dermatol. Surg.* **22**, 217–226 (2008).
- S. G. Gonzalez, M. Gill, and A. Halpern, *Reflectance Confocal Microscopy of Cutaneous Tumors—An Atlas with Clinical, Dermoscopic and Histological Correlations*, Informa Healthcare, London (2008).
- S. Nori, F. Rius-Díaz, J. Cuevas, M. Goldgeier, P. Jaen, A. Torres, and S. González, "Sensitivity and specificity of reflectance-mode confocal microscopy for *in vivo* diagnosis of basal cell carcinoma: A multicenter study," *J. Am. Acad. Dermatol.* **51**, 923–930 (2004).
- G. Pellacani, P. Guitera, C. Longo, M. Avramidis, S. Seidenari, and S. Menzies, "The impact of *in vivo* reflectance confocal microscopy for the diagnostic accuracy of melanoma and equivocal melanocytic lesions," *J. Invest. Dermatol.* **127**(12), 2759–2765 (2007).
- P. Guitera, G. Pellacani, C. Longo, S. Seidenari, M. Avramidis, and S. W. Menzies, "*In vivo* reflectance confocal microscopy enhances secondary evaluation of melanocytic lesions," *J. Invest. Dermatol.* **129**, 131–138 (2009).
- P. Guitera, G. Pellacani, K. A. Crotty, R. A. Scolyer, L. L. Li, S. Bassoli, M. Vinceti, H. Rabinovitz, C. Longo, and S. W. Menzies, "The impact of *in vivo* reflectance confocal microscopy on the diagnostic accuracy of lentigo maligna and equivocal pigmented and nonpigmented macules of the face," *J. Invest. Dermatol.* (2010).
- P. Calzavara-Pinton, C. Longo, M. Venturini, R. Sala, and G. Pellacani, "Reflectance confocal microscopy for *in vivo* skin imaging," *Photochem. photobiol.* **84**(6), 1421–1430 (2008).
- A. Gerger, S. Koller, W. Weger, E. Richtig, H. Kerl, H. Samonigg, P. Krippel, and J. Smolle, "Sensitivity and specificity of confocal laser-scanning microscopy for *in vivo* diagnosis of malignant skin tumors," *Cancer* **107**(1), 193–200 (2006).
- M. Vestergaard, P. Macaskill, P. Holt, and S. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting," *Bri. J. Dermatol.* **159**(3), 669–676 (2008).
- S. Rajpara, A. Botello, J. Townsend, and A. Ormerod, "Systematic review of dermoscopy and digital dermoscopy/artificial intelligence for the diagnosis of melanoma," *Bri. J. Dermatol.* **161**(3), 591–604 (2009).
- J. Braga, A. Scope, I. Klaz, P. Mecca, S. Gonzalez, S. H. Rabinovitz, and A. Marghoob, "The significance of reflectance confocal microscopy in the assessment of solitary pink skin lesions," *J. Am. Acad. Dermatol.* **61**(2), 230–241 (2009).
- S. Koller, M. Wiltgen, V. Ahlgrimm-Siess, W. Weger, R. Hofmann-Wellenhof, E. Richtig, J. Smolle, and A. Gerger, "*In vivo* reflectance confocal microscopy: automated diagnostic image analysis of melanocytic skin tumours," *J. Eur. Acad. Dermatol. Venereol.* (2010).
- S. Kurugol, J. G. Dy, M. Rajadhyaksha, and D. Brooks, "Detection of the dermis/epidermis boundary in reflectance confocal images using multi-scale classifier with adaptive texture features," in *Proc. IEEE Int. Symposium on Biomed. Imaging: From Nano to Macro*, pp. 492–495 (2008).
- S. Kurugol, J. G. Dy, M. Rajadhyaksha, and D. Brooks, "Localizing the dermis/epidermis boundary in reflectance confocal microscopy images with a hybrid classification algorithm," in *Proc. IEEE Int. Symposium on Biomed. Imaging: From Nano to Macro*, pp. 1322–1325 (2009).
- T. Randen and J. H. Husoy, "Filtering for texture classification: a comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 291–310 (1999).
- M. Tuceryan and A. Jain, "Texture analysis," in *Handbook of Pattern Recog. and Comp. Vis.* pp. 235–276, World Scientific, Singapore (1993).
- G. Doretto, D. Cremers, P. Favaro, and S. Soatto, "Dynamic texture segmentation," in *ICCV '03: Proc. Ninth IEEE Int. Conf. on Computer Vision*, p. 1236 (2003).
- S. Paoletti, A. L. Juloski, G. Ferrari-Trecate, and R. Vidal, "Identification of hybrid systems: A tutorial," *Eur. J. Control* **13**(2), 242–260 (2007).
- N. Ozay, M. Sznajder, and O. Camps, "Sequential sparsification for change detection," in *IEEE Conf. Computer Vision and Pattern Recognition, 2008*, pp. 1–6, (2008).
- M. Rajadhyaksha, S. Gonzalez, J. Zavislan, R. Anderson, and R. H. Webb, "*In vivo* confocal scanning laser microscopy of human skin II," *J. Invest. Dermatol.* **113**, 293–303 (1999).
- R. Gonzales and R. Woods, *Digital Image Processing*, Prentice Hall, Englewood cliffs NJ (2002).
- M. Wiltgen, A. Gerger, C. Wagner, and J. Smolle, "Automatic identification of diagnostic significant regions in confocal laser scanning microscopy," *Methods Inf Med.* **47**(1), 14–25 (2008).
- R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE* **67**(5), 786–804 (1979).
- A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(11), 1186–1191 (1993).
- D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am.* **4**(12), 2379–2394 (1987).
- L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.* **5**, 1205–1224 (2004).
- V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York (1995).
- V. Vural, G. Fung, B. Krishnapuram, J. G. Dy, and B. Rao, "Using local dependencies within batches to improve large margin classifiers," *J. Mach. Learn. Res.* **10**, 183–206 (2009).
- G. E. Costin and V. J. Hearing, "Human skin pigmentation: melanocytes modulate skin color in response to stress," *FASEB J.* **21**(4), 976–994 (2007).
- A. Scope, I. Klaz, and J. Casagrande, Memorial Sloan-Kettering Cancer Center, New York, NY, Private communication (2008).
- N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines (and other kernel-based learning methods)*, Cambridge University Press, Cambridge, UK (2000).
- M. Maseali, G. Fung, and J. Dy, "From transformation-based dimensionality reduction to feature selection," in *Int. Conf. on Machine Learning*, Citeseer (2010).
- B. Huang and D. S. Gareau, "Toward automated detection of malignant melanoma," *Proc. SPIE* **7169**, 71690X (2009).