Validation Study of Automated Dermal/Epidermal Junction Localization Algorithm in Reflectance Confocal Microscopy Images of Skin

Sila Kurugol*[a], Milind Rajadhyaksha[b], Jennifer G. Dy[a], Dana H. Brooks[a]

[a]Electrical and Comp. Eng., Northeastern University, 360 Huntington Av., Boston, MA;
[b]Dermatology Service, Memorial Sloan Kettering Cancer Cnt., 160 East 53rd St., New York, NY

ABSTRACT

Reflectance confocal microscopy (RCM) has seen increasing clinical application for noninvasive diagnosis of skin cancer. Identifying the location of the dermal-epidermal junction (DEJ) in the image stacks is key for effective clinical imaging. For example, one clinical imaging procedure acquires a dense stack of 0.5x0.5mm FOV images and then, after manual determination of DEJ depth, collects a 5x5mm mosaic at that depth for diagnosis. However, especially in lightly pigmented skin, RCM images have low contrast at the DEJ which makes repeatable, objective visual identification challenging. We have previously published proof of concept for an automated algorithm for DEJ detection in both highly- and lightly-pigmented skin types based on sequential feature segmentation and classification. In lightly-pigmented skin the change of skin texture with depth was detected by the algorithm and used to locate the DEJ. Here we report on further validation of our algorithm on a more extensive collection of 24 image stacks (15 fair skin, 9 dark skin). We compare algorithm performance against classification by three clinical experts. We also evaluate inter-expert consistency among the experts. The average correlation across experts was 0.81 for lightly pigmented skin, indicating the difficulty of the problem. The algorithm achieved epidermis/dermis misclassification rates smaller than 10% (based on 25x25 mm tiles) and average distance from the expert labeled boundaries of ~6.4 μm for fair skin and ~5.3 μm for dark skin, well within average cell size and less than 2x the instrument resolution in the optical axis.

Keywords: confocal reflectance microscopy, image analysis, skin, classification.

1. INTRODUCTION

One of the most common cancer types is skin cancer. Every year, in the US alone, about 3.6 million new cases of skin cancers are diagnosed[1-2]. Skin cancer screen in clinic is performed with a visual examination by naked eye and with a dermoscope[3-4]. Biopsy and histology is performed when an abnormal skin region is located during a visual exam. Biopsies are invasive, painful, destroy the site and leave a scar. Studies show that around 80% of biopsies return negative results.

Noninvasive imaging of skin for cancer screening and diagnosis with reflectance confocal microscopy (RCM) has been studied and reported in previous studies[5-9]. Epidermis and superficial dermis layers below the surface of the skin can be imaged with RCM. 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 optical sectioning of 1-3 μm and lateral resolution of 0.5-1.0 μm, which is comparable to that of conventional pathology. Sensitivity and specificity of detecting skin cancer with RCM reported in recent studies shows that RCM is advancing toward clinical utility for early noninvasive screening and diagnosis of skin cancers in real time while minimizing the need for biopsies[8,10].

*sila@bwh.harvard.edu; www.ece.neu.edu/~kurugol.s

Photonic Therapeutics and Diagnostics VIII, edited by Nikiforos Kollias, et al., Proc. of SPIE Vol. 8207, 820702

RCM acquisition is performed by sequentially capturing optical sections at increasing depths in skin. Horizontal slices acquired at each depth are recorded as a stack of images (see Fig.1) The point spread function and hence optical sectioning, resolution and contrast degrade with depth, due to increasing aberrations and scattering. Thus, detection of certain morphologic features remains challenging. Also, unlike pathological sections that are oriented perpendicular to the skin surface and are stained purple and pink, reflectance confocal images are oriented parallel (en face) and appear in grayscale (unstained). Thus, the appearance of RCM images is quite different from that of pathology. Visual evaluation of the features requires substantial training. Thus, computer automated image analysis tools to assist clinicians with evaluation and training could lead to adoption and a wider clinical utility of this otherwise attractive technology.

So far there are few publications on computer automated processing of RCM images to automatically identify quantitative features[11-14]. An example of a clinically important feature is the dermis/epidermis junction (DEJ), which is the 3 dimensional irregular surface separating the superficial epidermis from the underlying deeper dermis. The DEJ is clinically and pathologically important to examine, because cancers often originate and later spread from this location. Therefore, evaluation of the DEJ is important for early diagnosis.

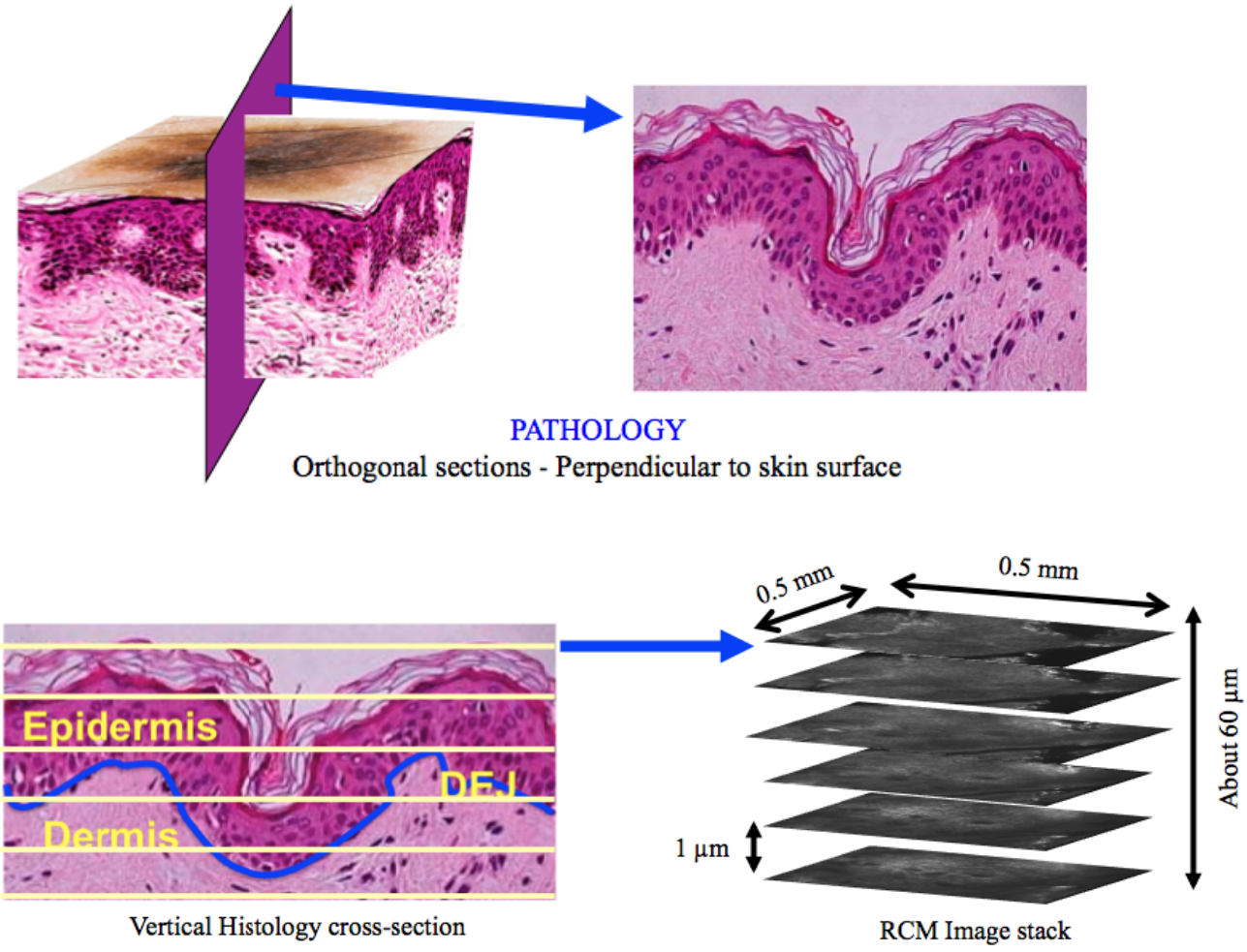


Figure 1. Comparison of pathology and histology with RCM. The top panel shows skin tissue on the left and a vertical histology cross-section on the right. The bottom panel shows vertical histology cross-section on the left with the blue curve drawn to indicate the location of the DEJ. The yellow lines indicate the horizontal slices imaged with RCM. A 3D RCM image stack is shown on the right.

Computer-automated image analysis may assist clinicians with the detection of the DEJ (and other morphologic features). However, in RCM images, the DEJ, like many other such features, is marked by optically subtle changes and features and is difficult to detect, with particular difficulty for lightly pigmented skin types where RCM contrast at the DEJ is poor (see Figs. 2 and 3). Additional challenges for automated-image analysis of RCM stacks from skin include heterogeneity of skin tissue, high inter- and intra-subject variability and low optical contrast. To overcome these challenges, we proposed a hybrid segmentation/classification algorithm for DE junction localization in lightly pigmented skin types[15-17]. This approach was a combination of two algorithms: First algorithm is the sequential image segmentation algorithm that partitioned the image sequences in depth (z) direction into homogenous groups using the dynamics of image features. Then, the second one, the machine learning–based locally smooth classification algorithm labeled these groups as epidermis and dermis regions sequentially. Both algorithms used a set of textural features calculated form the en face images.

Recently, we extended this algorithm to locate the DE junction in dark skin[18], in which strong backscatter from the melanin pigment causes the basal layer right above the DE junction to appear bright and with high contrast and was easier to detect compared to DE junction in fair skin stacks. In dark skin RCM stacks, the algorithm found the appropriate peak of the smoothed average intensity depth profile of an image region centered at position (x,y). To do so, we used 2-D texture features computed for each tile corresponding to a peak in intensity depth profile of that tile and automatically selected the right peak corresponding to basal cells by a texture similarity based analysis.

We also proposed a skin type detection algorithm[18], which decided the skin type of a given RCM stack based on existence of basal layer. After skin type detection, the appropriate DEJ localization method for either fair or dark skin was applied to that stack.

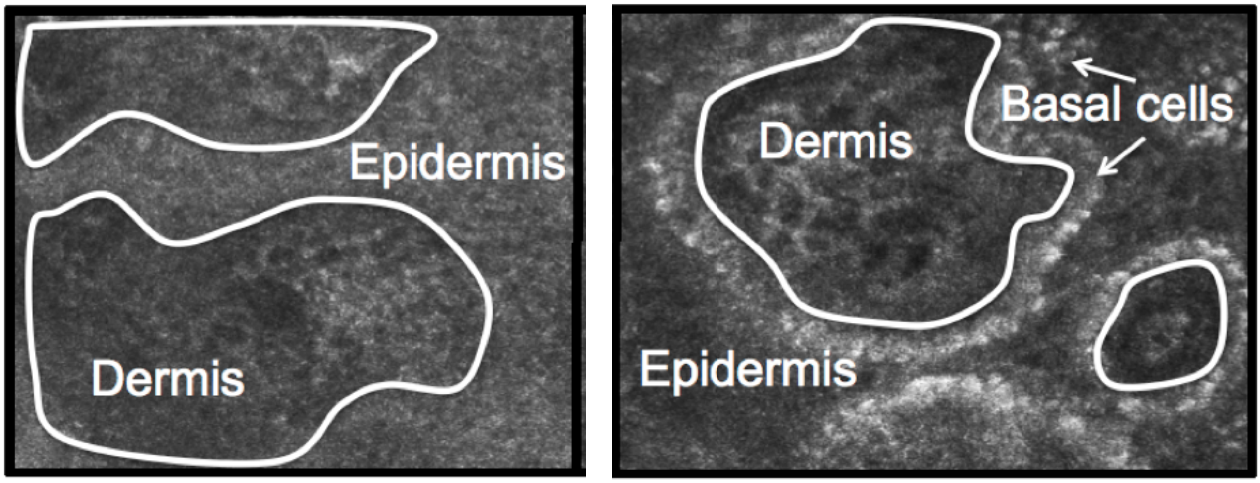


Figure 1. The left and right panels show two slices from an RCM stack from fair skin (on the left) and dark skin (on the right) respectively. The white boundary drawn is the DE junction.

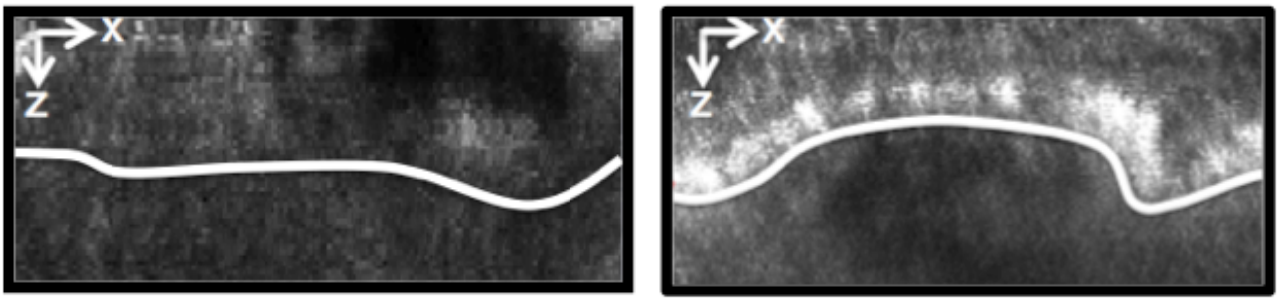


Figure 2. The left and right panels show two vertical slices from an RCM stack from fair skin (on the left) and dark skin (on the right) respectively. The white boundary drawn is the DE junction.

Here we report on the results of a validation study that we performed over 24 RCM stacks (9 dark skin, 15 fair skin) of normal skin to compare the DEJ detected by the algorithms with the DEJ manually labeled by experts. The results indicated that the algorithm localized DE junction for those test stacks with average errors of 5.32±4.27 μm for the DEJ in dark skin and 6.82±5.44 μm for epidermis boundary and 6.04±5.07μm for dermis boundary in fair skin.

2. METHODS

2.1 Data acquisition and preprocessing

Imaging was performed on healthy volunteers with a VivaScope 1500™ (Lucid Inc., Rochester NY, operating at 785 nm wavelength). A stack of horizontal images was captured starting at a position below the skin surface. The resolution was 0.5 μm in lateral direction and about 3 μm in the axial direction. The field-of-view was 500μm. After capturing each image, the en-face optical section was translated 1 μm deeper along the optical axis into the tissue from the starting position below the skin surface to a sufficient depth to be in the dermis. The clinician performing the imaging selected both starting and end points as arbitrary positions in epidermis and dermis respectively. After acquisition of a confocal stack, the stack of 8-bit tiff images were loaded into Matlab software for automated processing to locate the DEJ. The image stack was first converted into a volume matrix and the automated preprocessing algorithm was applied.

The images were not aligned and there was shift in lateral directions (x and y) from one image to the next due to patient movement during acquisition. To correct for this shift a standard stack registration algorithm was applied.

2.2 Automated DEJ Detection Algorithm

Skin Type Detection: For a given RCM stack, we first applied the automated skin type detection algorithm we proposed in previous work[18], which determined whether the given stack was dark skin (pigmented skin) or fair skin (very lightly pigmented skin) (See Figs. 2 and 3). To determine the skin type, one obvious useful feature is the presence of very bright basal cells in dark skin, which are not present in fair skin. Therefore these basal cells need to be searched for within the given stack; if they are present, we can conclude that the stack is from dark skin.

After skin type detection, according to the detected skin type, the DEJ detection algorithm for fair or dark skin was applied to the given stack. Both DEJ detection algorithms operated on tiles, i.e. small square regions that are large enough to include a few cells; hence the first step was partitioning of the stack into tiles. To detect the boundaries of epidermis and dermis layers (i.e. the DEJ), both fair and dark skin algorithms used the dynamics of skin layer appearance with depth information.

Automated DEJ detection for dark skin: In dark skin, the basal layer has bright basal cells including highly reflective melanin pigment. The DEJ is located at the lower boundary of the basal layer, separating the basal layer from the underlying dermis. To detect DEJ in dark skin, due to existence of strong intensity contrast at basal layer, intensity information with depth was used. However, this strong peak in intensity at the basal layer was not consistent across the stack. Some tiles had multiple strong peaks due to bright appearing epidermis region or deep dermal collagen fibers. For those tile a texture based basal layer detection algorithm was applied to select the peak including basal cells. This algorithm was proposed and explained in detail in our previous work[18].

Automated DEJ detection for fair skin: For fair skin types, due to low amount of melanin pigment, the basal cells do not appear bright in RCM stacks. Therefore, for these stacks the DE junction detection task is harder due to the lack of contrast and strong features, as well as heterogeneity of skin tissue. In fair skin RCM stacks, instead of detecting a strict DE junction, a transition zone was detected. This transition zone has upper boundary, (i.e. lower boundary for epidermis layer) and has lower boundary (i.e. upper boundary for dermis layer). The DE junction is located in between these two boundary surfaces. To detect these dermis and epidermis boundaries, due to lack of contrast, instead of intensity information, we utilized the texture dynamics of skin tissue in depth direction, as proposed in our previous work[17].

The algorithm used 2D texture features calculated for a z-stack of tiles to discriminate the textural differences between different skin layers. The set of features used in automated skin type detection and DE junction detection algorithms were the same set of features used in previous work[17-18]. From each tile, we extracted this same set of 170 texture features including gray level co-occurrence matrix features (contrast, energy, correlation and homogeneity), statistical metrics (mean, variance, skewness and kurtosis), features from a wavelet decomposition[19], log-Gabor features and radial spectral features. From these 170 features, we selected the most discriminative and least redundant subset of features with an automated feature selection algorithm over a training set of manually labeled stacks[17,20].

The dynamics of the skin texture in depth was represented by the multivariate feature sequence in depth calculated for a tile stack. This feature sequence was used to partition the stack into homogenous segments in z (depth). To do so, a model of skin layer dynamics was fitted to these features from the z-stack of tiles. Then, those tile segments were classified as epidermis, dermis, or transitional DEJ region using texture features. The classifier was trained on an RCM stack where the DE junction was manually labeled and applied to new RCM stacks to automatically locate the DE junction.

3. RESULTS

We applied the DEJ detection algorithm on 24 RCM stacks (15 fair skin stacks, 9 dark skin stacks) from our database. For each stack, we had ground truth (expert labeling) available. We compared the boundaries located by the algorithm with the ground truth. Table 1 shows the mean and standard deviation of the distances between expert labeled DEJs and the automatically located DEJs in 9 dark skin RCM stacks. The DEJs found by the algorithm (dotted red) and the DEJs marked by the expert (green) are compared for two sample vertical cross sections (x-z) and (y-z) from the first two stacks from Table 1 in Fig. 3 and 4.

Surface plot of the DEJs automatically found by the algorithm are shown in 3D in comparison to expert labeled DEJ for the first three stacks from Table 1 in Fig 5. The surface itself indicates the resultant DEJ of the algorithm and the color map indicates the distance from the expert labeled DEJ (error).

In Fig. 6, the resultant automatically located epidermis (shaded with red) and dermis regions (shaded with blue) are shown for various axial RCM slices at various depths imaged parallel to the skin surface for RCM stack 1.

Table 1. Table shows the mean and standard deviation of the distances between expert labeled DE junctions and the automatically located DE junctions in 9 RCM stacks from dark skin.

RCM Stack	Mean ± σ (μm)
1	4.47±3.13
2	3.31±3.11
3	2.90±3.02
4	2.88±3.22
5	8.43±6.46
6	9.26±8.44
7	9.93±4.56
8	2.61±3.45
9	4.16±3.04
Mean ±σ (μm)	**5.32±4.27**

The results of the DE junction detection algorithm for fair skin were reported in our previous work for 4 RCM stacks. Here we report the results for 15 fair skin RCM stacks. We compared the epidermis and dermis boundaries located by the algorithm with the ground truth. The DEJ is located in between these boundaries. However, at some regions, where

wrinkles are present, DEJ location is not calculated, therefore those regions are excluded from the results. Table 1 shows the mean and standard deviation of the distances between expert labeled epidermis and dermis boundary surfaces and the automatically located surfaces in 15 fair skin stacks. The last column shows the average error over all stacks.

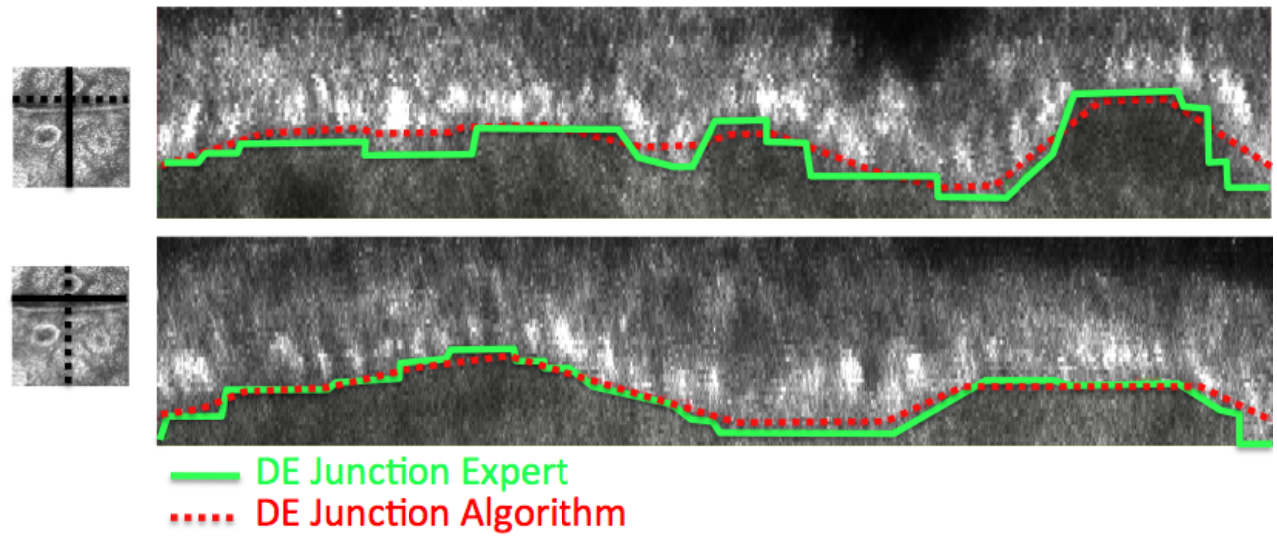


Figure 3. The upper and lower panels on the right compare the DE junction found by the algorithm (dotted red) with the one marked by the expert (green) for two sample vertical cross sections (x-z) and (y-z) from the RCM stack 1-3. The solid lines in the left figures indicate the vertical slice location on a sample horizontal slice. Note that the expert marks the DE junction not on the vertical slices but on horizontal slices.

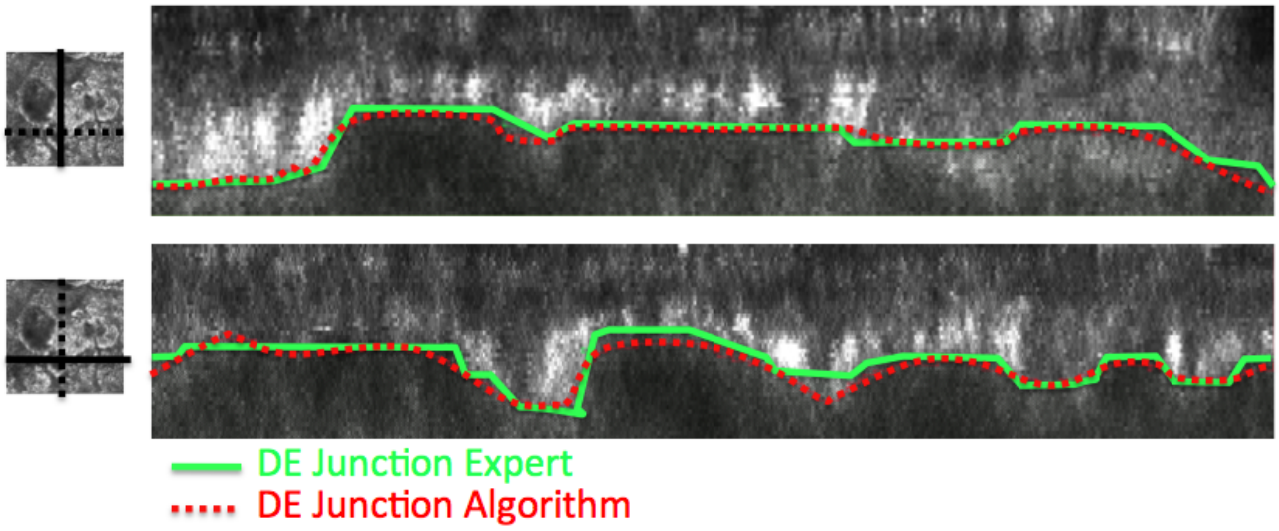


Figure 4. Similar figure as Fig. 3 for RCM stack 2

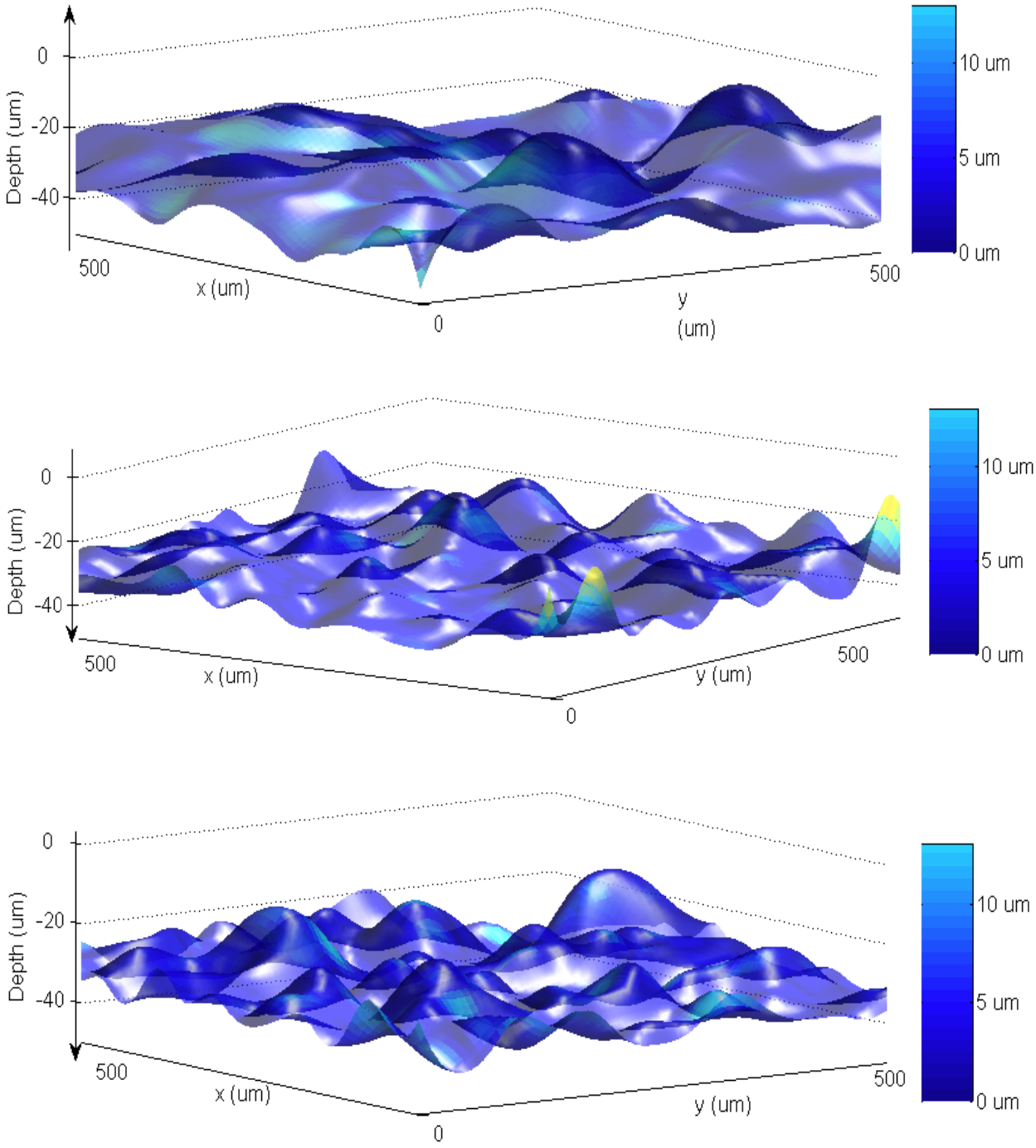


Figure 5. Surface plots of the DE junctions automatically found by the algorithm are shown in 3D in comparison to expert labeled DE junctions for RCM stacks 1 to 3 from dark skin. The surface itself indicates the resultant DE junction of the algorithm and the color map indicates the distance from the expert labeled DE junction (error).

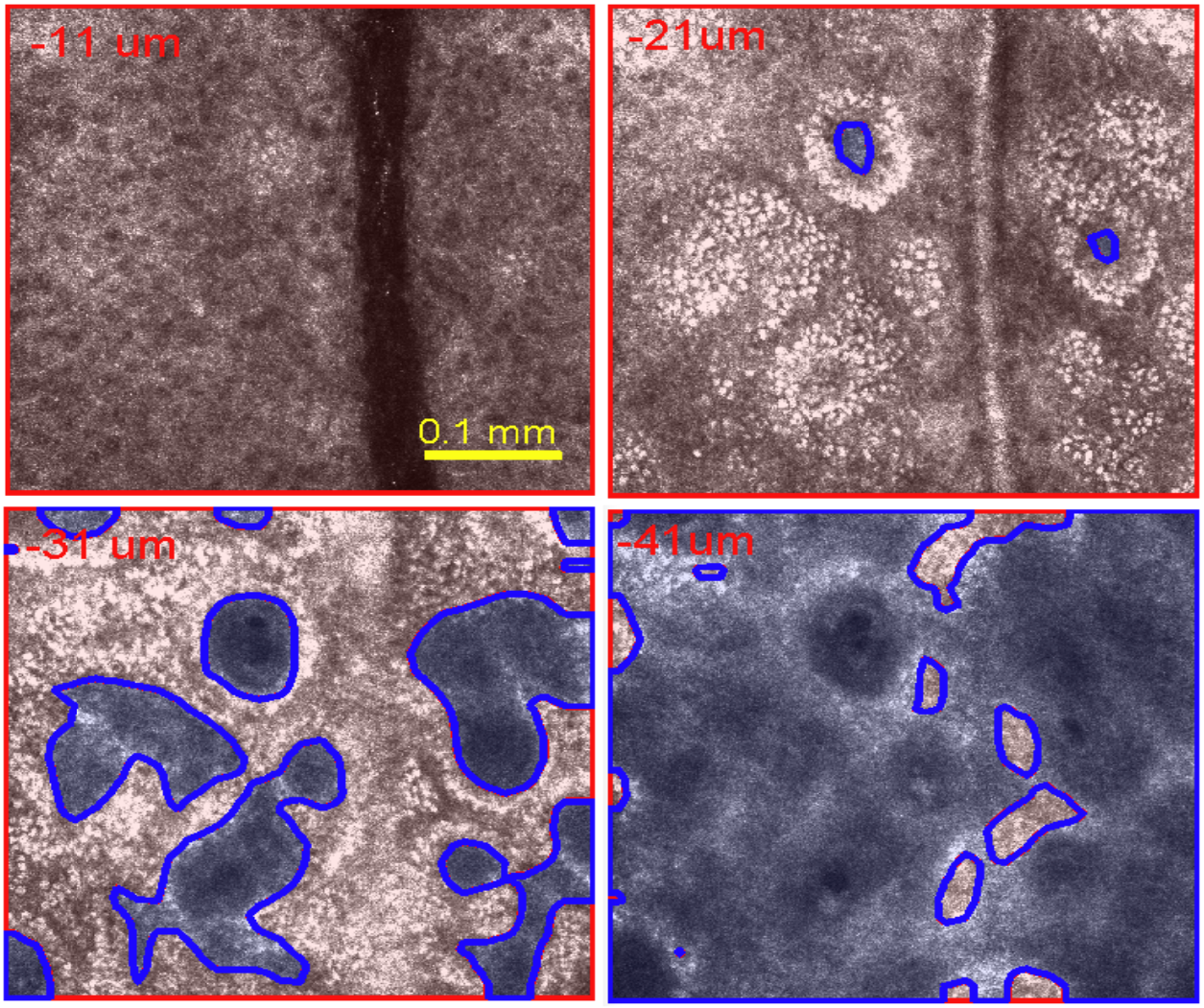
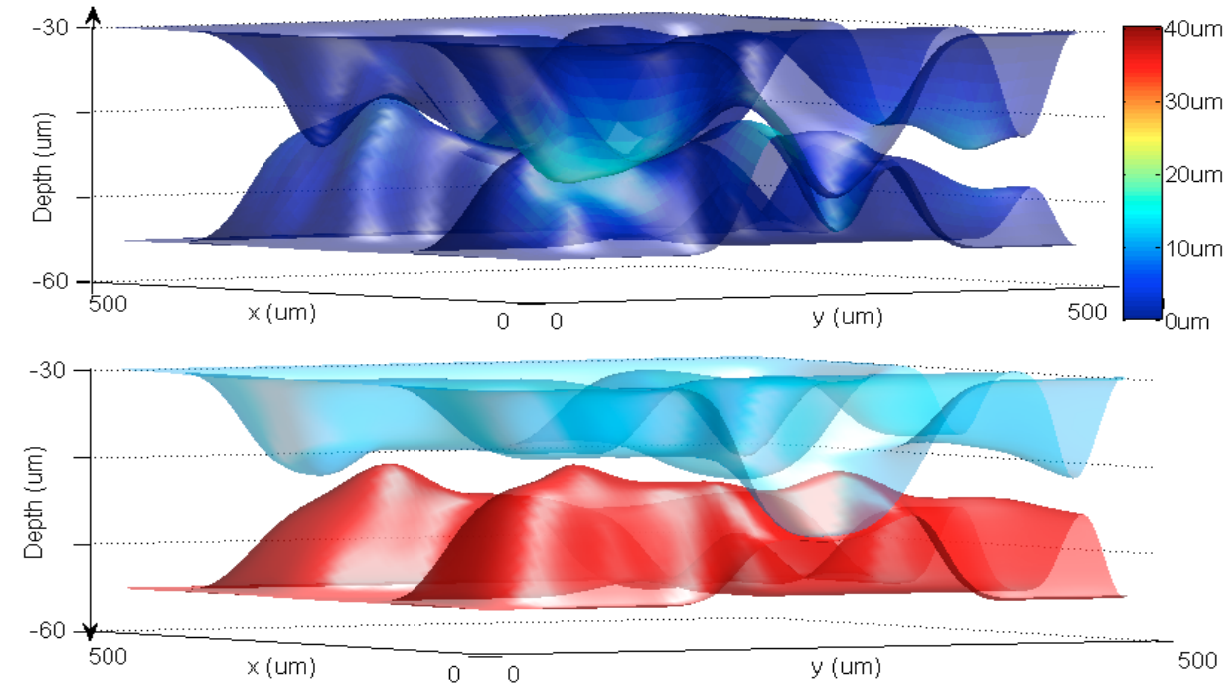


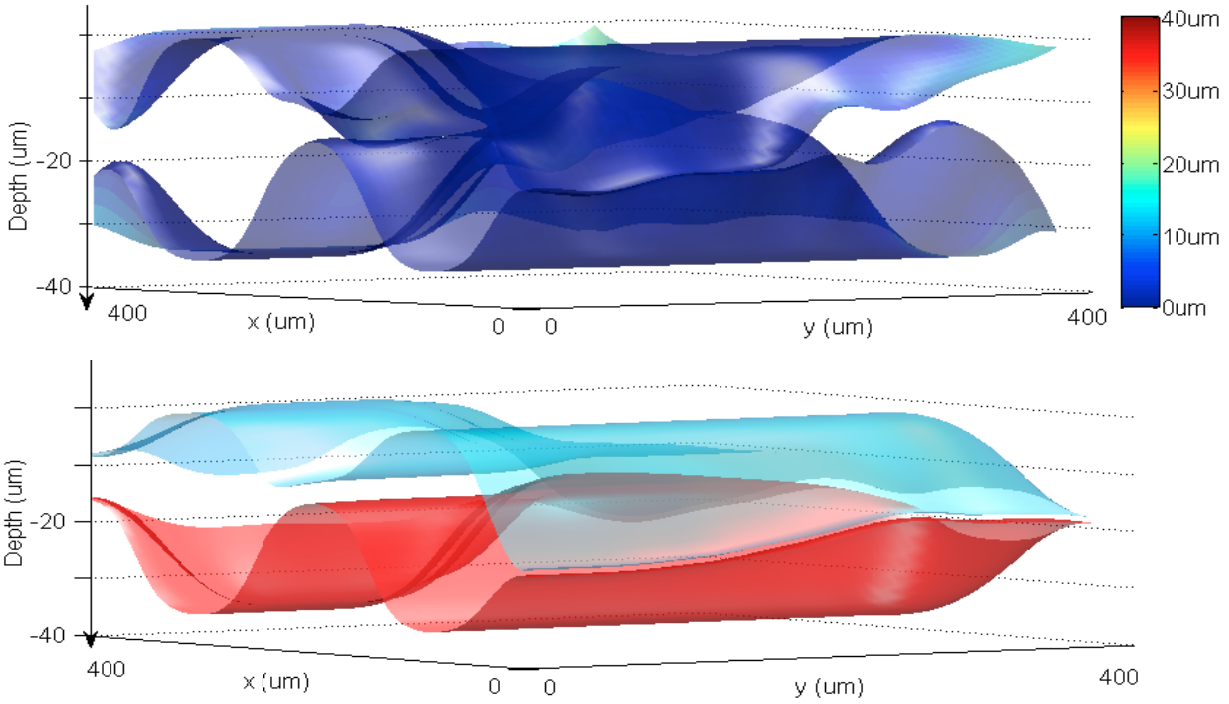
Figure 6. The resultant automatically located epidermis (red) and dermis (blue) regions are shown on the axial RCM slices at various depths imaged parallel to the skin surface for RCM stack 1 from dark skin. The numbers on top left corner indicates the depth in reference to the most superior axial slice of the image stack

Table 2. The mean and standard deviation of the distances between expert labeled epidermis and dermis boundaries and the automatically located epidermis and dermis boundaries in 15 RCM stacks from fair skin are reported. The last column of the table shows the mean over all 15 stacks.

RCM Stack no	1	2	3	4	5	6	7	8
Epidermis Mean±σ (μm)	7.89 ± 7.37	7.77± 6.35	5.89± 5.19	5.46± 3.92	11.02±9.62	5.75± 3.70	3.68±2.89	6.63± 4.12
Dermis Mean±σ(μm)	5.33 ± 5.11	7.29± 7.40	5.19± 5.12	4.24± 3.40	9.73±7.92	5.38± 4.11	2.24± 1.77	5.89±3.04
RCM Stack no	9	10	11	12	13	14	15	**Mean±σ(μm)**
Epidermis Mean±σ(μm)	7.12± 3.70	8.08± 6.54	8.46±6.53	5.44± 4.77	5.60± 5.47	7.90±6.15	5.73± 5.28	**6.82± 5.44**
Dermis Mean±σ(μm)	7.29±3.99	5.53± 6.13	6.94±6.61	5.83± 4.82	6.14± 5.33	8.81± 6.65	4.83±4.79	**6.04±5.07**



(a) RCM Stack 4



(b) RCM Stack 7

Figure 6. Surface plot of the epidermis boundary and the dermis boundary in 3D in comparison to the expert labeled boundaries of RCM stack 4 and 7 from Table 2. Top blue (bottom red) surfaces show the expert labeled epidermis (dermis) boundary The colored surfaces indicate the resultant boundaries of the algorithm. The color maps indicate the distance from the expert labeled boundary. The axes are is in micrometers. 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.

We also evaluated the inter-expert consistency among the experts. We asked all of our three expert clinicians to label the same fair skin RCM stack. On that stack, the average correlation calculated across experts was 0.81 for lightly pigmented skin, indicating the difficulty of the problem.

4. CONCLUSION AND FUTURE WORK

In this work, we performed a validation study for the algorithms we proposed earlier to to detect whether the RCM stack is from light or dark skin type and to then locate the DEJ surface in RCM image stacks using the DEJ detection algorithm for either dark or fair skin types. The skin type detection algorithm classified the stacks including reflective basal cell as dark skin. The dark skin DEJ detection algorithm first detected the peaks in the mean intensity profiles for each tile and then selected the peaks that corresponded to the basal cells. After locating the basal cells, the lower boundary of the basal cells corresponding to the DEJ was found and the DEJ surface was constructed. The fair skin algorithm used texture featured to first partition z-stacks of tiles into homogenous segments corresponding to skin texture changes with depth and then detected the epidermis and dermis layers sequentially with a locally smooth classifier. The results show that the DEJ algorithm for dark skin type resulted in reasonable performance with average distance from the ground truth DEJ surface around 5.32μm. Similar results for DEJ detection algorithm for fair skin types resulted in epidermis/dermis misclassification rates smaller than 10% and average distance from the expert labeled boundaries around 6.4 μm.

ACKNOWLEDGEMENTS

We thank Dr. Alon Scope, Dr. Juliana Casagrande and Dr. Itay Klaz for providing us the manual segmentations of DEJ in RCM images. This research was supported by the Center for Integrative Biomedical Computing (CIBC) under grants from the National Center for Research Resources (5P41RR012553-14) and the National Institute of General Medical Sciences (8 P41 GM103545-14) from the National Institutes of Health. Support for M.R. was provided by NIH grant no. R01EB012466.

REFERENCES

[1] Gloster, H. and Brodland, D. G., "The epidemiology of skin cancer," British Journal of Dermatology 22, 217-226 (2008).
[2] American *Cancer* Society, "Cancer Facts *&* Figures," (2011).
[3] Vestergaard, M., Macaskill, P., Holt, P., and Menzies, S., "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting," British Journal of Dermatology 159(3), 669-676 (2008).
[4] Rajpara, S., Botello, A., Townend, J., and Ormerod, A., "Systematic review of dermoscopy and digital dermoscopy/artificial intelligence for the diagnosis of melanoma," British Journal of Dermatology 161(3), 591-604 (2009).
[5] Gonzalez, S. G., Gill, M., and Halpern, A., [Reflectance Confocal Microscopy of Cutaneous Tumors: An Atlas with Clinical, Dermoscopic and Histological Correlations], Informa Healthcare, London (2008).
[6] Nori, S., Rius-Díaz, F., Cuevas, J., Goldgeier, M., Jaen, P., Torres, A. and González, S.,"Sensitivity and specificity of reflectance-mode confocal microscopy for in vivo diagnosis of basal cell carcinoma: A multicenter study," J. American Academy of Dermatol 51, 923{930 (2004).
[7] Pellacani, G., Guitera, P., Longo, C., Avramidis, M., Seidenari, S. and Menzies, S., "The impact of in vivo reflectance confocal microscopy for the diagnostic accuracy of melanoma and equivocal melanocytic lesions," J. Invest. Dermatol 147(14), 2759-65 (2007).

[8] Guitera, P., Pellacani, G., Longo, C., Seidenari, S., Avramidis, M. and Menzies, S.W., cIn Vivo Reflectance Confocal Microscopy Enhances Secondary Evaluation of Melanocytic Lesions," J. Invest. Dermatol 149, 131-138 (2009).
[9] Guitera, P., Pellacani, G., Crotty, K.A., Scolyer, R.A., Li, L.X., Bassoli, S., Vinceti, M., Rabinovitz, H., Longo, C. and Menzies, S.W., "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).
[10] Calzavara-Pinton, P., Longo, C., Venturini, M., Sala, R. and Pellacani, G., "Reflectance Confocal Microscopy for In Vivo Skin Imaging," Photochemistry and photobiology 84(6), 1421-1430 (2008).
[11] Gerger, A., Koller, S., Weger, W., Richtig, E., Kerl, H., Samonigg, H., Krippl, P. and Smolle, J., "Sensitivity and specificity of confocal laser-scanning microscopy for in vivo diagnosis of malignant skin tumors," Cancer 107(1), 193-200 (2006).
[12] Gareau, D.S., Hennessy R., Wan E., Pellacani G., Jacques S.L., "Automated detection of malignant features in confocal microscopy on superficial spreading melanoma versus nevi," J. Biomedical Optics, 15(6), (2010).
[13] Koller, S., Wiltgen, M., Ahlgrimm-Siess, V., Weger, W., Hofmann-Wellenhof, R., Richtig, E., Smolle and J., Gerger, A., "In vivo reflectance confocal microscopy: automated diagnostic image analysis of melanocytic skin tumours," Journal of the European Academy of Dermatology and Venereology: JEADV, (2010).
[14] Wiltgen, M., Gerger, A., Wagner, C., Smolle, J., "Automatic Identification of Diagnostic Significant Regions in Confocal Laser Scanning Microscopy of Melanocytic Skin Tumors," Methods Inf. Med., 47(1), pp. 14-25 (2008).
[15] Kurugol, S., Dy, J.G., Brooks, D. H. and Rajadhyaksha, M., "Localizing the dermis/epidermis boundary in reflectance confocal microscopy images with a hybrid classification algorithm," Proc. of IEEE Int. Symposium on Biomed. Imaging: From Nano to Macro," pp. 1322-1325 (2009).
[16] Kurugol, S., Dy, J.G., Brooks, D. H. and Rajadhyaksha, M., "Detection of the dermis/epidermis boundary in reflectance confocal images using multi-scale classifier with adaptive texture features," Proc. Of IEEE Int. Symposium on Biomed. Imaging: From Nano to Macro, pp. 492-495 (2008).
[17] Kurugol, S., Dy, J.G., Brooks, D. H. and Rajadhyaksha, M., "A pilot study of semi-automated localization of the dermal/epidermal junction in reflectance confocal images of skin," J. Biomedical Optics 16 (3), (2011).
[18] S. Kurugol, J. G. Dy, M. Rajadhyaksha, K. W. Gossage, J. Weissman, Dana H. Brooks. "Semi-automated Algorithm for Localization of Dermal/ Epidermal Junction in Reflectance Confocal Microscopy Images of Human Skin", In Proc. SPIE Photonics West, (2011).
[19] Gonzales, R., and Woods, R., [Digital Image Processing], Prentice Hall (2002).
[20] Randen, T. and Husoy, J.H., "Filtering for texture classification: a comparative study," IEEE Trans. Pattern Anal. Mach. Intel., 21, 291-310 (1999).