Automatic Localization of Skin Layers in Reflectance Confocal Microscopy

Eduardo Somoza, Gabriela Oana Cula(✉), Catherine Correa, and Julie B. Hirsch

Johnson and Johnson Consumer Companies, Inc.
Skillman, NJ, USA
{esomoza1,gcula,ccorrea1,jhirsch}@its.jnj.com

Abstract. Reflectance Confocal Microscopy (RCM) is a noninvasive imaging tool used in clinical dermatology and skin research, allowing real time visualization of skin structural features at different depths at a resolution comparable to that of conventional histology [1]. Currently, RCM is used to generate a rich skin image stack (about 60 to 100 images per scan) which is visually inspected by experts, a process that is tedious, time consuming and exclusively qualitative. Based on the observation that each of the skin images in the stack can be characterized as a texture, we propose a quantitative approach for automatically classifying the images in the RCM stack, as belonging to the different skin layers: stratum corneum, stratum granulosum, stratum spinosum, stratum basale, and the papillary dermis. A reduced set of images in the stack are used to generate a library of representative texture features named textons. This library is employed to characterize all the images in the stack with a corresponding texton histogram. The stack is ultimately separated into 5 different sets of images, each corresponding to different skin layers, exhibiting good correlation with expert grading. The performance of the method is tested against three RCM stacks and we generate promising classification results. The proposed method is especially valuable considering the currently scarce landscape of quantitative solutions for RCM imaging.

Keywords: Reflectance confocal microscopy · Image stacks · Skin texture · Textons · Clustering · Dimensionality reduction · Classification · Image recognition

1 Introduction

Reflectance Confocal Microscopy (RCM) is employed as a non-invasive imaging tool in clinical dermatology and skin research, as well in the evaluation of changes due to application of products in the cosmetics industry [1,2]. RCM allows for real time visualization of structural features at different skin depths at a resolution comparable to that of conventional histology [1].

The foundation of RCM imaging is based on the generation of endogenous contrast due to differences in the degree of light reflected from different structures [3]. In RCM, the surface of interest is illuminated by a laser light source and the reflected light is collected through a pinhole aperture where the photons of light encounter a photodetector [3]. Structures with highly reflective surfaces will have a high degree of

A. Campilho and M. Kamel (Eds.): ICIAR 2014, Part II, LNCS 8815, pp. 141–150, 2014.
DOI: 10.1007/978-3-319-11755-3_16

light dispersion and as a result, will produce a high intensity signal when illuminated by the laser light source; in contrast, structures that have poorly reflective surfaces will produce a low intensity signal. Some of the structures found in skin with high reflectance capabilities in comparison to surrounding structures are melanosomes, cytoplasmic granules, cellular organelles, and keratin-containing structures [1]. Different skin layers contain different reflective structures that can be used to distinguish each layer from one another during image acquisition with RCM.

RCM's depth of penetration allows for the visualization of all four layers of the epidermis (stratum corneum, stratum granulosum, stratum spinosum, stratum basale) and portions of the papillary dermis (uppermost layer of the dermis), depending on the thickness of the skin, as illustrated in Figure 1. In thinner regions of the skin, it is possible to reach the superficial reticular dermis (a lower level of the dermis) [4]. RCM is capable of providing a lateral resolution of approximately 1 μm, an axial resolution of 3-5 μm, and a penetration depth of 150-300 μm depending on the instrumentation and image acquisition parameters [3,4,5].

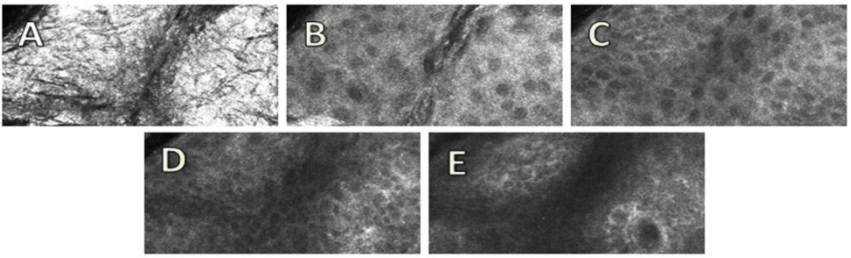

Fig. 1. Typical images within a normal skin RCM image stack. RCM's depth of penetration allows for the visualization of all four layers of the epidermis: stratum corneum (A), stratum granulosum (B), stratum spinosum (C), stratum basale (D) and portions of the papillary dermis, the uppermost layer of the dermis (E), depending on the thickness of the skin. Images in this figure were used as the training set for our method.

The penetration depth of RCM is highly dependent on the power of the laser component of the system: a higher power laser will allow for acquisition of images in the deeper layers of skin but at the same time, there will be a sacrifice in resolution and potential damage to skin structures [3]. Depending on the step size used for image acquisition, the resulting RCM image stack can contain anywhere from tens to hundreds of images. Once the RCM stack is acquired, typically, expert-based visual assessment of each image for characteristic structural and cellular features that distinguish adjacent skin layers needs to be performed to classify each image [5]. This visual assessment is time consuming when dealing with multiple RCM stacks with hundreds of images of different anatomical sites of skin. While visually investigating skin RCM stacks of images, we have observed a smooth change in the textural properties of the images as a function of depth. Therefore we propose a method based on texture characterization of different skin layers to classify RCM images into their respective layers as a function of depth.

1.1 Prior Work

Reflectance of light on the skin surface has been studied in relation to the appearance of skin texture. Different illumination parameters (e.g. direction and angle of light) lead to different visual perceptions of the appearance of skin texture due to differences in the reflectance of light. These changes in appearance are problematic for the identification and classification of different skin features: a single skin feature might be identified as a different skin feature depending on the illumination parameters. To gain independence from changes in appearance of texture due to illumination, texton-based classification systems have been employed [6,7,8]. Textons are texture representations of characteristic structural features found on a surface of interest.

By compiling an extensive vocabulary of textons for visually related surfaces of interest, classification and separation of images of these surfaces can be accomplished through texton-based texture representation. In the dermatological application of RCM, these closely related surfaces are different layers of the skin encountered during image acquisition. The RCM images of different layers of skin change smoothly from one acquisition step to the next, and in this work we generate texture representations for each layer based on a library of textons, which are texture representations for characteristic structural features of the inner layers of skin. Using this texton-based approach, we are automatically able to classify RCM images as one of the five skin layers encountered during RCM image acquisition. After the classification of individual images, we automatically separate adjacent skin layers. We then correlate the results from the automatic classification and separation of the RCM images to the classification and separation of images by expert visual assessment. To our knowledge, this is the first attempt in trying to classify RCM images of skin layers using a texton-based approach, while previous attempts utilized an intensity-based approach [9,10].

2 Methods

2.1 Acquisition of Representative RCM Image Stack

A representative image stack of normal skin was acquired with the Vivascope 1500 (Lucid Technologies, Rochester, NY, USA) using 785 nm laser illumination. The acquisition of the representative RCM stack was performed on the volar forearm region of a healthy adult male because of the ease of placement of the Vivascope 1500 confocal head on the skin surface, leading to a reduction in imaging artifacts due to movement. The image stack was collected at a step size of 1 μm, starting at the stratum corneum and ending at the uppermost layers of the papillary dermis. The dimensions of the images were 1000 by 1000 pixels. All pixels of the skin images in the representative stack were normalized to a normal distribution with parameters (μ=0, σ=1). Images were classified into five layers of interest through visual assessment by an RCM expert: stratum corneum (SC), stratum granulosum (SG), stratum spinosum (SS), stratum basale (SB), and the papillary dermis (D).

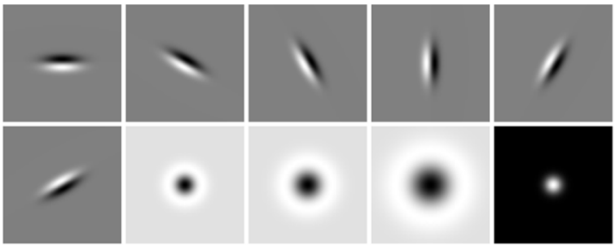

Fig. 2. The selection of 10 filters we employ in our method from the Leung-Malik filter bank [7]. The filters we select are six oriented derivative of Gaussians, three Laplacian of Gaussian derivative filter and one Gaussian filter.

2.2 Generation of Texton Filter Response Space

A representative image from each of the layers of interest is selected to create the training set, illustrated in Fig.1, needed to generate the texton filter response space. To ensure the training set consisted of images that exhibit characteristic structural features of each of the five skin layers, a 250 by 500 pixel subsection is chosen as part of the training set. To capture local orientation patterns of characteristic structural features, 10 filters from the Leung Malik (LM) filter bank are selected to create the filter set used to generate the texton filter response space [7]. The filters in the LM filter bank have been previously employed to generate texture representations of surfaces because of their ability to capture structural parameters, such as local orientation parameters, that contribute to texture representation for classification [6,7,8]. To decrease the likelihood of incorporating noise generated during image acquisition into the texton filter response space, the dimensions of the filters are matched to the dimensions of a keratinocyte (a predominant structural feature in skin), which are about 25 by 25 pixels [4] . The texton filter response space is generated by the convolution of the individual images in the training set with each of the filters in the filter set. The image windows of size 25x25 to which the filter set is applied to are normalized to a normal distribution with parameters (μ=0, σ=1).

2.3 Generation of Texton Library

The dimensionality of the filter response space generated was ten: each dimension corresponding to one of the filters. We use principal component analysis (PCA) to reduce the dimensionality of the filter response space. We find that 90% of the variability in the filter response space generated by filtering the representative images from each of the layers of interest could be accounted for by the first three principal components. The filter response space is re-projected onto these three principal components, reducing the dimensionality of the space from ten to three dimensions. To generate the texton library, K means clustering is performed on the reduced filter response space, resulting in 15 representative textons (cluster centers in the filter response space). K means clustering is also performed to generate a lower number of textons (10) and a higher number of textons (25), but we have determined empirically that 15 textons sufficiently characterize the structural features apparent in the RCM images.

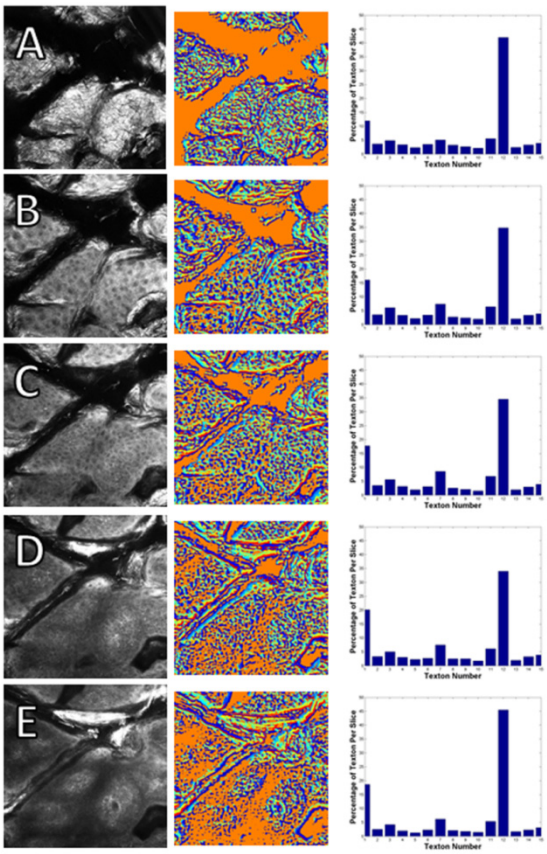

Fig. 3. Classification of representative images based on the texton library. A representative texton histogram was generated for each image showing the percentage of textons used in the classification. From Top to Bottom, Left to Right: **(A)** Stratum Corneum original image, classified image, and representative texton histogram. **(B)** Stratum Granulosum original image, classified image, and representative texton histogram. **(C)** Stratum Spinosum original image, classified image, and representative texton histogram. **(D)** Stratum Basale original image, classified image, and representative texton histogram. **(E)** Papillary Dermis original image, classified image, and representative texton histogram.

2.4 Classification of Images and Layer Separation

The same filter set is applied to all images within the RCM stack and the resulting filter response vectors are re-projected onto the same three principal components determined during the generation of the texton library. Each image pixel is classified as one of the 15 representative textons based on the shortest square Euclidean distance between the projected filter response and the textons in the library. The resulting classified image is the texture representation of the image based on the texton library. A representative 15-dimensional texton histogram is generated for each of the images and is employed to separate the RCM stack into the 5 representative skin layers.

To reduce the dimensionality of the collective texton histograms corresponding to the images within the RCM stack, PCA is performed one more time. It is determined that 90% of the variability in this universal eigenspace is accounted for by the first three principal components. The texton histograms are re-projected onto these three principal components, reducing the dimensionality of the histogram space from 15 to 3 dimensions. In this reduced 3D eigenspace, K-Means clustering with K=5 is performed to automatically separate the RCM stack into the five representative layers. Consequently, each texton histogram is classified to the closest center based on square Euclidean distance. The results of the automatic separation are compared to the separation of the RCM stack based on visual assessment by the RCM expert. To assess the consistency of the proposed method, this classification process is repeated with two other visually classified RCM image stacks of the volar forearm region from two different healthy adults with normal skin.

3 Results

3.1 Selection of Training Set and Filter Set

To generate the representative texton library, a training set of images and a filter set are selected. Five representative images of the SC, SG, SS, SB and D selected to be included in the training set. To ensure the training set captured the characteristic structural features of each layer as identified by RCM experts in the literature [5], a 250 by 500 pixel subsection is selected. The result was a training set (Figure 1) encompassing a majority of the characteristic structural features of the representative skin layers, but also excluding non characteristic structural features such as noisy patches.

The filters of the filter set are chosen from the LM filter bank: a filter bank that has been used in previous works concerning texture representations of textures [6,7,8]. Ten filters (Figure 2) from the LM filter bank are chosen for their ability to encode patterns of local orientations of characteristic structural features in the RCM images.

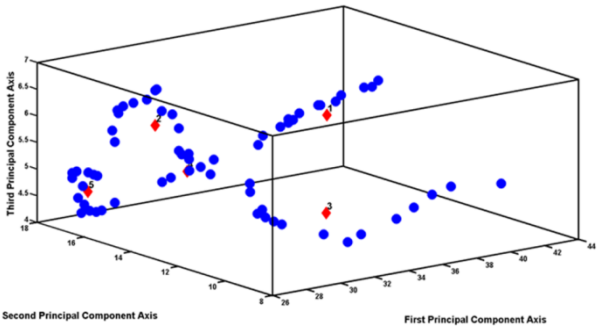

Fig. 4. K Means clustering of texton histogram space. The cluster centers (represented by a red diamond marker), as determined by K Means, correspond to a different skin layer classification within the imaging stack. The individual images in the stack are represented by blue circle markers.

3.2 Classification of RCM Images Using Texton Library

Each of the training images are filtered with the filters selected from the LM filter set, and a 10 dimensional filter response space is created. Each dimension corresponds to one of the filter responses obtained by convolving of the images with one of the filters in the filter set. Applying PCA on the texton response space, three principal components are identified and the space is re-projected onto these components. K Means clustering is performed on the reduced filter response space to generate the 15 textons in the library. The images in the representative RCM stack are labeled using the resulting library of 15 textons and each image is further represented by a texton histogram. Figure 3 shows the results from the texton classification of the five representative images of the original training set.

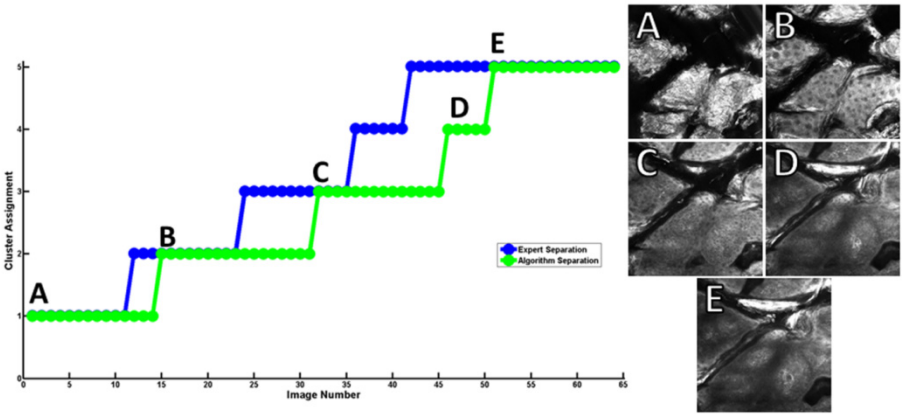

Fig. 5. Graphical representation of automatic separation of representative stack based on results from K-Means clustering of texton histogram space versus expert labeled images. Each cluster assignment, marked as 1 to 5 on the Y-axis corresponds to a different skin layer. From lowest to highest assignment, we find the following layers: Stratum Corneum, Stratum Granulosum, Stratum Spinosum, Stratum Basale, and Papillary Dermis. The images on the right (labeled A-E) are the training RCM images and they are found by the proposed method as belonging to the layer the expert grader assigns them to. The correlation coefficient corresponding to assessing the correlation between the automatic evaluation and that of the expert grader is 0.916.

3.3 Separation of Representative RCM Stack

The dimensionality of the texton histogram space is further reduced by applying PCA, and a three dimensional eigenspace is generated. Within this newly reduced texton histogram eigenspace, K Means clustering is performed to generate five representative centers corresponding to each of the five layers represented in the training set. The results of K Means clustering of the texton histogram space are shown in Figure 4, where a clear separation of the skin layers is evident.

Each of the five centers corresponds to a different layer of the skin. Each of the representative texton histograms of the images in the RCM stack are classified as one

of these representative centers. The result of the classification leads to an automatic separation of the RCM stack into the five different representative skin layers. The resulting separation is a smooth transition between adjacent skin layers as a function of depth, indicating each layer is represented by a set of characteristic structural features (Figure 5). The results from the automatic separation were compared to the results from the separation based on the visual assessment of a RCM expert (Figure 5). Figure 5 shows some misalignment between the automatic classifications of the RCM stack versus the expert's visual assessment. This misalignment is addressed in more detail in Section 4.

To check the consistency of the algorithm in its ability to classify images and separate different layers of the skin within any RCM stack, the same texton library is used to classify and separate three additional RCM stacks. These RCM stacks are images of the volar forearm of three healthy adults, acquired and processed the same way as the representative RCM stack used to generate the texton library images. The results of the automatic classification and separation are compared to the visual assessment by an RCM expert (Figure 6). The results are very promising as the corresponding correlation coefficients vary between 0.84 and 0.95.

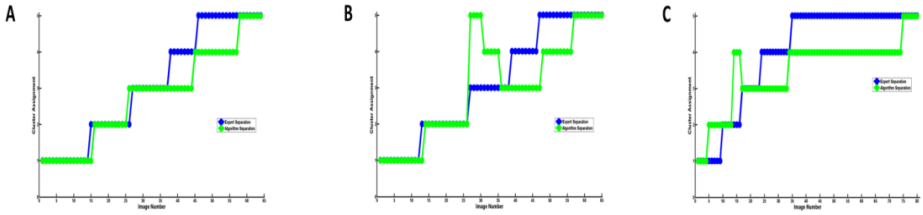

Fig. 6. We test the method on three extra RCM image stacks. Each graph corresponds to a different stack. The green line is the output from our method while the blue line corresponds to expert grading. We correlate the method based skin layer evaluation with that assessed by the expert and we obtain correlation coefficients 0.954, 0.844 and 0.865, respectively.

4 Discussion and Conclusions

We have identified the need for automatic image classification in normal skin RCM image stacks and we made the observation that skin RCM images at each step in the image stack have a repetitive nature and could be represented as textures. To capture the repetitive nature of the appearance in these images, we use the concept of texton, introduced in [11] as the up to second-order statistic of textural images that is relevant for texture discrimination at pre-attentive level of human perception. As in [6], we have generated a representative texton library for the characteristic structural features of the inner most layers of skin captured during image acquisition with RCM. With these textons, we have been able to create texture representations of each image in an RCM stack. With these texture representations, we classified each image into one of five skin layers: stratum corneum, stratum granulosum, stratum spinosum, stratum basale, and the papillary dermis. From the classification of each of the images, we automatically separate adjacent layers in the RCM stack.

The texton-based classification methodology proposed here is not only limited to being used for the classification and separation of RCM skin images of normal skin. This methodology can be used to study different diseased states of skin such as eczema, squameous cell carcinoma, and psoriasis by expanding our current texton library to include texture representations of characteristic structural features present in these diseases. With an expanded texton library, we can generate texture representations of the layers of the diseased skin and see if these representations change after medical treatment. This same approach could be used in the cosmetic industry when testing the efficacy of different moisturizing products. The texton library could further be expanded to include representations of structural skin features found in infant, adolescent, and elderly populations, hence studying the effects of developmental and aging processes on the skin.

We have compared the automatic classification and separation of three RCM skin image stacks with the results of visual assessments performed by an RCM expert. The results of these correlations are excellent with the correlation coefficients vary from 0.84 to 0.95. However, we observe a slight misalignment between the automatic and visual assessment results, which could be explained by the fact that the methodology proposed here relies only on texture characteristics of structural features in each layer to generate its classifications, while the RCM expert relies on high-level multiple factors for his assessment: the relative depth of the image within the RCM stack, the intensity of signal from low/high reflective surfaces, and cellular features (e.g. shape, size).

Currently, the textons are defined at pixel level and they encode local texture information that is a small part to what the expert perceives as different cells present in skin layers. Next steps include expanding the concept of textons to that of a more complex structural element that captures high level information relevant for the expert grader. Specifically, we plan to define the concept of "hierarchical texton" that will be able to go beyond representing the texture at pixel level, but to capture the relative positioning of textons up to the level of new textons being capable to represent features at cellular level. Moreover, next steps also include developing a supervised classification model capable of learning associations between expert-labeled skin features and computationally based texture features, ensuring that the expert information is encompassed in the computational model to maximize correlation between automatic classification and the expert grader outcome.

References

1. Gonzalez, S., Gilaberte-Clazada, Y.: In Vivo Reflectance-mode Confocal Microscopy in Clinical Dermatology and Cosmetology. International Journal of Cosmetic Science **30**(1), 1–17 (2008)
2. Hofmann-Wellenhof, R., Pellacani, G., Malvehy, H., Soyer, P. (eds.) Reflectance Confocal Microscopy for Skin Diseases. Springer (2012)
3. Rajadhyaksha, M., Gonzalez, S., Zavislan, J.M., Anderson, R.R., Webb, R.H.: In Vivo Confocal Laser Microscopy of Human Skin II: Advances in Instrumentation and Comparison with Histology. Journal of Investigative Dermatology **113**(3), 293–303 (1999)

4. Sanchez-Mateos, J.L.S., Rajadhyaksha, M.: Optical Fundamentals of Reflectance Confocal Microscopy. Monografias de Dermatologia **24**(2), 1–3 (2011)
5. Sanchez, V.P., Gonzalez, S.: Normal Skin. Monografias de Dermatologia **24**(2), 1–3 (2011)
6. Cula, O.G., Dana, K.J.: Skin Texture Modeling. International Journal of Computer Vision **62**(1/2), 97–119 (2005)
7. Leung, T., Malik, J.: Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons. International Journal of Computer Vision **43**(1), 29–44 (2001)
8. Varma, M., Zisserman, A.: Classifying images of materials: achieving viewpoint and illumination independence. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 255–271. Springer, Heidelberg (2002)
9. Kurugol, S., Dy, J.G., Rajadhyaksha, M., Gossage, K.W., Weissmann, J., Brooks, B.H.: Semi-automated Algorithm for Localization of Dermal/Epidermal Junction in Reflectance Confocal Microscopy Images of Human Skin. In: Proceedings of SPIE 7904, Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XVIII, 79041A (2011)
10. Kurugol, S., Rajadhyaksha, M., Dy, J.G., Brooks, D.H. : Validation Study of Automated Dermal/Epidermal Junction Localization Algorithm in Reflectance Confocal Microscopy Images of Skin. In: Proceedings of SPIE 8207, Photonic Therapeutics and Diagnostics VIII, 820702 (2012)
11. Julesz, B.: Textons, the Elements of Texture Perception, and their Interactions. Nature **290**(1), 91–97 (1981)