UNSUPERVISED WRINKLE DETECTION IN REFLECTANCE CONFOCAL MICROSCOPY IMAGES OF THE HUMAN SKIN

J. Sourati[1], D. H. Brooks[1], J. G. Dy[1], E. Ataer-Cansizoglu[1], D. Erdogmus[1], M. Rajadhyaksha[2]

[1] Northeastern University, Boston, MA 02115, USA
[2] Memorial Sloan-Kettering Cancer Center, Dermatology Service, New York, NY, USA

ABSTRACT

Reflectance confocal microscopy (RCM) is a non-invasive and *in-vivo* imaging modality, which can take images from different depths of the human skin. A challenging problem is to detect a clinically important subsurface section of the skin, the Dermis/Epidermis junction, in RCM images. This is a tough problem because of the huge variation of texture and intensity features across both intersubject and intrasubject tissues. On the other hand, there's almost no wrinkle-free part of the skin. This well-known phenomenon can be used as a histological clue for guessing the probability of being Dermis or Epidermis in the neighboring regions. In this paper, we develop a two-step wrinkle detector for RCM images. By analyzing the results on different RCM images, we conclude it has high sensitivity and specificity, but a relatively lower Jaccard index.

Index Terms— Reflectance confocal microscopy, Dermis-Epidermis junction, wrinkle detection

1. INTRODUCTION

Reflectance confocal microscopy (RCM) is an *in-vivo* imaging modality which takes advantage of different reflectance factors of various materials by radiating a light beam and spatially filtering the beam to measure the reflectance intensity of the focused voxel. It uses optical sectioning to give a complete image of the volume. Skin imaging is one of the main applications of RCM which can be done in real-time and does not need the extraction of sample tissue as in biopsy, which is invasive, time-consuming, and destructive for cells structure.

On the other hand, classification of the human skin is of specific interest for the clinicians, since many fatal skin diseases start growing from the boundary of the first two superficial layers, Epidermis and Dermis. Generally, this type of classification is very tough because of large texture and intensity feature variations across intersubject and intrasubject tissues. So we need to somehow add histological information about different structures in the skin in order to get access to more reliable clues about the Dermis-Epidermis Junction (DEJ) location. Wrinkles are well-known skin phenomena that can provide us with such additional information. It's been shown that wrinkles have influence on the shape of DEJ in its surrounding regions [2] (see Figure 1).

In [1] Kurugol et. al employed a local texture-based classification method by using Locally Smooth SVM to detect the DEJ. However, that method excludes the wrinkles in the first step and does the classification for the rest of the voxels. Here, as a first step towards involving histological information, we are interested in detecting the wrinkles as helpful structures that can be detected as a preprocessing step for DEJ localization in RCM images.

To the authors' best knowledge there has not been any attempts to automatically detect wrinkles in RCM skin images. Usually wrinkles exhibit themselves as valleys in three-dimensional visualization of pixel intensity elevation. From this perspective, they are similar to vessels and one might expect to find them by using standard ridge/valley detection algorithms [5]. In detecting curvilinear structures, such as Keratin intermediate Filaments [3], typically ridges are enhanced using methods such as Frangi and Satu filters, both of which analyze the eigenvectors of the Hessian matrix in each pixel to determine whether the pixel is on the ridge or not. In our application eigenvector analysis alone cannot be sufficient, since, as discussed above, mild or intense textures inside the wrinkles makes this type of ridge detection and enhancement filters to fail. Overall, although they provide satisfactory results in some wrinkle regions, generic vessel extraction filters will lead to failure in many others in our experience. The major challenge is the frequent

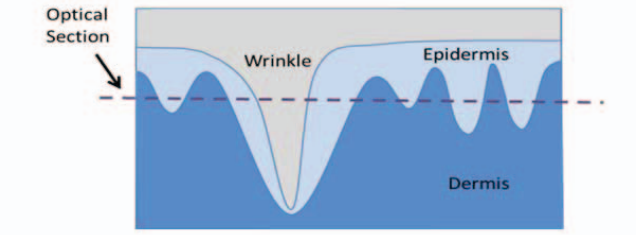


Figure 1. Topography of a wrinkle and the optical section of RCM which is parallel to the skin surface

This work has been supported by NSF and NIH.

occurrence of highly textured regions within some parts of the wrinkles due to another histological phenomenon called *horny plug* [2]. This structure makes wrinkles appear very similar to non-wrinkle parts of the RCM image. Furthermore, the wrinkle boundaries are less defined in the edge distribution when compared to tubular structures such as neurons, airways, or blood vessels. Moreover, in many cases we cannot distinguish between wrinkles and ordinary regions except by looking at the continuity of neighboring textures.

For these reasons, using a single-step global algorithm is challenging; instead, we will develop a two-step algorithm, which combines both global (the first step) and local (the second step) information to perform wrinkle classification.

2. METHOD

Our method has two steps: Initialization and Finalization. In the former one we provide initial regions which highly probably belong to a wrinkled area, then we grow this region to finalize the estimated wrinkle. In both steps we use texture features to exclude regions with high frequency intensity fluctuations from the detected area. For this purpose, we employ the energy of wavelet packet coefficients [4] at level three[1], denoted by $\mathbf{wp}_3(x,y)$ for pixel at coordinate (x,y). As Figure 1 shows, the edges of wrinkles are actually the most superficial layer of the skin, which has a relatively high reflectance factor and thus appear brightly in RCM images. Therefore, if we start from one point within the wrinkle and grow the initial region based on texture features, we expect that the selected region will remain inside the wrinkle and will not leak out because of this natural boundary. Keeping this in mind, we develop the algorithm such that it starts by selecting some initial regions with high probability of belonging to wrinkles and then try to grow this region in order to cover rest of the wrinkle details.

2.1. Initialization

The primary feature of wrinkled regions, despite occasional high intensity variations, is their darkness. We choose to select a subset of wrinkled low-intensity regions as the initialization since it's easier and less risky.

Let us denote the input image by $I(x,y)$. For the rest of the paper we denote W_t as the set of superpixels yielded from applying Watershed-based segmentation on the thresholded image I'_t defined later; $Wl_t \subseteq W_t$ as the subset of superpixels, which are larger than a pre-specified threshold a_{th} (empirically determined as 150 pixels in our case); P_t and Pl_t as the sets of all pixels fallen in W_t and Wl_t respectively and P_w as the set of pixels within actual wrinkles. Finally s_A stands for index of superpixels in an arbitrary set A and $|\ |$ is an operator over sets that returns number of members; since each s_A is a set of pixels $|s_A|$ is the number of pixels in it. First of all we smooth the floor of wrinkle valleys by applying a shifted step function over the image:

$$I'_t(x,y) = u_t.I(x,y) = \begin{cases} I(x,y), & I(x,y) > t \\ 0, & o.w. \end{cases} \quad (1)$$

If the shift value t is selected properly, then we will have a relatively smoother valley in the wrinkles, while few number of non-wrinkle pixels would be shrunk to zero. This leads to having a piecewise connected floor of the valleys, which can also be viewed as connected sources. Hence if we use a watershed-based segmentation [6] we will get huge superpixels for the connected sources and smaller ones in the textured regions. Then by simply picking out sufficiently large superpixels, we can generate the initialization area. Clearly choosing the value of t is very critical in the result of this step and consequently in the success or failure of the whole algorithm. We adjust it adaptively based on the local minima of the intensity histogram of the image.

Table 1 shows the initialization algorithm; the threshold is chosen among the histogram minima (the set M) that are lower than the intensity mean by minimizing the criterion fraction c_t shown in the table. As we go toward higher values of t bigger superpixels will be created but the probability of leakage into the non-wrinkle parts also increases; we also don't want the superpixels to be too small (shrinkage) because then our initialization region doesn't have sufficient coverage for feeding the finalization step. In order to prevent leakage, we keep texture energy of the selected region small by involving the energy of wavelet packet coefficients of level three (computed by means of built-in functions of MATLAB), while trying to moderate shrinkage by introducing the average area of the selected superpixels in the denominator of c_t.

If there are considerable pixels belonging to wrinkles there would be clear minima and maxima in the lower half of the histogram; in the case of extreme smoothness when no minima is found we divide the whole image into four

Table 1. Algorithm Initialization

h = smoothed histogram of $I(x,y)$ using Gaussian kernel
M = $\{m_1, m_2, \dots, m_N\}$,
where m_i is a local minimum of **h** less than E($I(x,y)$)
for t : 1→ N
  $W_t \leftarrow$ watershed($I'_t(x,y)$)
  $Wl_t \leftarrow \{s_{W_t}: |s_{W_t}| > a_{\text{th}}\}$
  $Pl_t \leftarrow \{(x,y): (x,y) \in Wl_t\}$
  $c_t = \sum_{(x,y)\in Pl_t} ||\ \mathbf{wp}_3\ (x,y)||_2 / \text{E}(|s_{W_t}|)$
end
$t^* = \underset{t}{\text{argmin}}\ c_t$
$W_{t^*} \leftarrow$ watershed($I'_{t^*}(x,y)$)
$Wl_{t^*} \leftarrow \{s_{W_{t^*}}: |s_{W_{t^*}}| > a_{\text{th}}\}$
$P_{\text{init}} = \{(x,y): (x,y) \in Wl_{t^*}\}$

[1] This is compatible with the current resolution of our database; for lower resolution we may find it better to change the level.

subimages and apply the same algorithm to each of the divisions which gives us four initializations; then the final result would be their union. Finally if we couldn't find any minima for any of the subimages we conclude that the input image doesn't contain any wrinkled region.

2.2. Finalization

We barely can find cellular or collagen textures within the wrinkles, as severe as in Epidermis or Dermis tissues. Based on this fact, we perform superpixel-wise region growing algorithm to expand the initialized region. Even if there are non-wrinkle regions with textural features as mild as the wrinkled parts, the boundaries will most probably prevent leakage to them because of their high reflectance property mentioned before.

Let W_0 and $W_{E(\mathbf{h})}$ be the set of superpixels yielded from applying Watershed on the original and the over-smoothed image $u_{E(\mathbf{h})}.I(x,y)$. To avoid using the whole superpixels in W_0 for region-growing algorithm which contains large number of irrelevant superpixels we only consider the ones that are fallen inside the giant superpixels of $W_{E(\mathbf{h})}$:

$$W_c = \{s_{W_0} : s_{W_0} \subseteq Wl_{E(\mathbf{h})}\}. \quad (2)$$

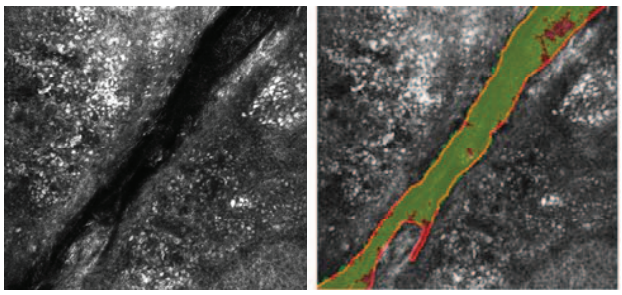
Since high frequency components are more discriminative, we exclude the coefficients corresponding to the extremely low frequencies from our texture feature vector and denote the resulting vectors by $\mathbf{wp}_{t,3}(x,y)$. Initialized region R can be characterized by its truncated textural energy mean:

$$\mathbf{e}(R) = \sum_{(x,y)\in Pl_{t^*}} \mathbf{wp}_{t,3}(x,y) / \sum_{s_R} |s_R|. \quad (3)$$

Every time we take the superpixels in W_c that are located on the boundaries of the new region and add the ones which have similar texture feature vectors with respect to $\mathbf{e}(R)$; more specifically experimental threshold of 5 is put over the Euclidean distance between $\mathbf{e}(R)$ and feature vector of the candidate superpixel. The iterations will be repeated until all candidates in W_c are tried or there is no candidate superpixel in the boundaries.

3. EXPERIMENTAL RESULTS

Our database includes 40 RCM images containing wrinkles from different sites of the body such as arm, knee, abdomen, forehead and back. The images are chosen so that they cover as many types of wrinkles' appearances as possible. They have resolution of 1µm and are from different depth of the skin. For evaluation purpose, we compute sensitivity (S_n) and specificity (S_p) for every image and report their mean and standard deviation (STD) (see Table 2). According to these measures, 88% of our wrinkle labels and 97% of our non-wrinkle labels were correctly classified. Wrinkles usually occupy a small portion of the whole image, therefore the number of true negatives (TN), considering wrinkle as the positive event, are always much bigger than number of falsely negatives. This causes specificity to be very high. In order to investigate the percentage of FN more precisely we also computed the Jaccard index ($= \frac{|\mathrm{TP}|}{|\mathrm{TP}|+|\mathrm{FN}|}$ where TP is number of true positives), which is the portion of the true wrinkle captured by the algorithm (detection rate). Table 2 reveals that the result is less than S_n as expected.



(a) (b)

Figure 2. (a) Original wrinkled skin image, (b) expert's label (red) and resulted segmentation of the algorithm (green)

Table 2. Evaluation of the algorithm's performance

	Mean	STD
Sensitivity	0.88	0.08
Specificity	0.97	0.02
Jaccard index	0.72	0.13
WJ index ($d_c = 15$)	0.9	0.15

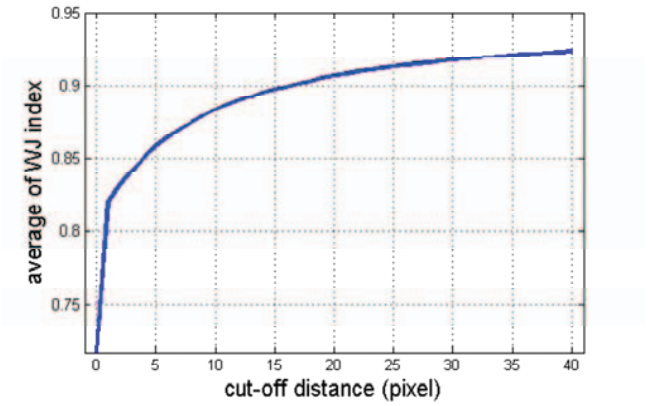


Figure 3. Average WJ index for different cut-off distances; $d_c = 0$ corresponds to the original Jaccard index.

Note that we aim to use the wrinkle labels as source for developing prior probability models for the surrounding regions—specifically regarding the location of the DEJ from the surface. In many cases undetected wrinkle parts are too close to the detected parts such that their surrounding regions appear very closely; thus missing a few pixels from a detected wrinkle region would have less impact on the eventual goal of DEJ localization than an entirely missed wrinkle.

Figure 2 shows a sample image from the dataset with the expert's labels and the results of the algorithm. It can be seen that part of the wrinkle is missed in the upper right side of the image; however it's in the neighbourhood of the true positives the algorithm generated. The existence of sufficient true positives in the vicinity should help form a sufficient prior about the location of DEJ in this region despite missed wrinkle pixels. We defined weighted Jaccard (WJ) index as below to take into account the distance between missed wrinkles and true labels:

$$WJ = \frac{|\mathrm{TP}|}{|\mathrm{TP}|+\sum_{(x,y)\in \mathrm{FN}} B_c(x,y)}, \quad (4)$$

where $B_c(x,y)$ is the Butterworth bandpass filter in the distance domain:

$$B_c(x,y) = 1 - \frac{1}{1+(d(x,y)/d_c)^{2N}}, \quad (5)$$

in which $d(x,y)$ is the shortest Euclidean distance between the pixel location (x,y) and true labels. d_c is called cut-off distance and is defined as a distance to which we assign 1/2. Using this filter we actually weigh down the missed wrinkles close to TP region based on their distance and emphasize distant ones by effectively counting pixels in FN which are further than d_c. Table.2 shows the result for d_c=15. In Figure 3, average WJ is plotted for different values of d_c. It can be seen that even for low values of d_c we have a big jump in the mean implying that there are many undetected labels which are highly close to TPs such that they won't leave big effects in the future prior specification.

4. DISCUSSION AND CONCLUSION

We performed binary segmentation of RCM images of human skin to decide whether a region belongs to a wrinkle or not. This is an important issue because the existence of wrinkles affects the localization of Dermis-Epidermis Junction – a clinically very important region to investigate for various skin disorders and diseases. This wrinkle detector can be used as a preprocessing step that gives the surrounding regions of a wrinkle a relatively higher prior of belonging to the Epidermis. This is going to be quite useful in cases that we're in risk of misclassifying the regions around the wrinkles as Dermis because of reasons such as low resolution or imaging's artifacts.

The results indicate that the algorithm is able to achieve high sensitivity and specificity in addition to a sufficiently large Jaccard and Weighted Jaccard index for our purposes. That purpose is to detect wrinkles and use these pixels in building a prior model for skin surface, based on which the Dermis-Epidermis Junction can be found in conjunction with RCM imagery in real time. While the algorithm cannot detect all wrinkle pixels due to discontinuous wrinkle parts and anomalous skin structures which exhibit appearances very similar to non-wrinkle regions, the results are sufficiently accurate to build on.

The initialization step is designed such that it wouldn't put any seed points in low-confidence regions. Thus if there are discontinuous wrinkles which are recognized as suspicious regions they might be missed due to the region growing algorithm not being able to penetrate into these regions in the second step. By this type of initialization we are actually doing a conservative segmentation since the potential clinical cost of misdetection is much more than missing wrinkles. In particular, the existence of a wrinkle-labeled pixel will reduce the probability of having the DEJ in its vicinity; DEJ is exactly where clinically important events, such as malignant tumors, occur. Consequently, inserting false wrinkle pixels into the remainder of the algorithm pipeline might have serious negative outcomes. For this reason, missing a wrinkle extension at the cost of weakening the prior is acceptable.

The WJ index that we introduced also indicates that most of the missed wrinkle pixels are located near the true positives. Therefore, in these regions, the algorithm would likely be able to form a sufficiently strong prior anyway.

Although the initial performance we obtained in this study is satisfactory to move to the next step of detecting the DEJ, the wrinkle classifier could be further improved by incorporating additional features such as vesselness features based on eigenstructure of the local Hessian matrices.

5. REFERENCES

[1] S. Kurugol, J.G. Dy, D.H.Brooks and M. Rajadhyaksha, "Pilot Study of Semi-automated Localization of the Dermial/Epidermal Junction in Reflectance Confocal Microscopy Images of Skin," *Journal of Biomedical Optics*, March 2011.

[2] J.L. Contet-Audonneau, C. Jeanmaire, and G. Pauly, "A Histological Study of Human Wrinkle Structure: Comparison between Sun-exposed areas of the Face, with or without Wrinkles, and Sun-protected Areas," *British Journal of Dermatology*, pp. 1038-1047, 1999.

[3] G. Herberich, R. Windoffer, R. Leube, T. Aach, "3D Segmentation of Keratin Intermediate Filaments in Confocal Laser Scanning Microscopy," *IEEE Engineering in Medicine and Biology Conference*, Boston, pp. 7751-7754, 2011.

[4] Mallat, S., *A Wavelet Tour of Signal Processing*, Academic Press, New York University, 1998.

[5] C. Kirbas and F. Quek, "A Review of Vessel Extraction Techniques and Algorithms," *ACM Computing Surveys*, pp. 81-121, 2004.

[6] F. Meyer, "Topographic Distance and Watershed Lines," *Signal Processing 38*, pp. 113:125, 1994.