Wavelet-based statistical classification of skin images acquired with reflectance confocal microscopy

Abdelghafour Halimi, Hadj Batatia, Jimmy Le Digabel, Gwendal Josse, Jean-Yves Tourneret

Wavelet-based statistical classification of skin images acquired with reflectance confocal microscopy

Abdelghafour Halimi,[1] Hadj Batatia,[1,*] Jimmy Le Digabel,[2] Gwendal Josse,[2] and Jean Yves Tourneret[1]

[1] *University of Toulouse, IRIT-INPT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France*
[2] *Centre de Recherche sur la Peau, Pierre Fabre Dermo-Cosmétique, 2 rue Viguerie, 31025 Toulouse Cedex 3, France*
[*] *hadj.batatia@inp-toulouse.fr*

Abstract: Detecting skin lentigo in reflectance confocal microscopy images is an important and challenging problem. This imaging modality has not yet been widely investigated for this problem and there are a few automatic processing techniques. They are mostly based on machine learning approaches and rely on numerous classical image features that lead to high computational costs given the very large resolution of these images. This paper presents a detection method with very low computational complexity that is able to identify the skin depth at which the lentigo can be detected. The proposed method performs multiresolution decomposition of the image obtained at each skin depth. The distribution of image pixels at a given depth can be approximated accurately by a generalized Gaussian distribution whose parameters depend on the decomposition scale, resulting in a very-low-dimension parameter space. SVM classifiers are then investigated to classify the scale parameter of this distribution allowing real-time detection of lentigo. The method is applied to 45 healthy and lentigo patients from a clinical study, where sensitivity of 81.4% and specificity of 83.3% are achieved. Our results show that lentigo is identifiable at depths between $50\mu m$ and $60\mu m$, corresponding to the average location of the the dermoepidermal junction. This result is in agreement with the clinical practices that characterize the lentigo by assessing the disorganization of the dermoepidermal junction.

OCIS codes: (100.0100) Image processing; (100.2960) Image analysis; (100.7410) Wavelets; (170.0170) Medical optics and biotechnology; (170.0180) Microscopy; (170.6935) Tissue characterization.

References and links

1. K. S. Nehal, D. Gareau, and M. Rajadhyaksha, "Skin imaging with reflectance confocal microscopy," Seminars in Cutaneous Medecine and Surgery **27**, 37–43 (2008).
2. R. Hofmann-Wellenhof, E. M. T. Wurm, V. Ahlgrimm-Siess, E. Richtig, S. Koller, J. Smolle, and A. Gerger, "Reflectance confocal microscopy state-of-art and research overview," Seminars in Cutaneous Medecine and Surgery **28**, 172–179 (2009).
3. M. Rajadhyaksha, A. Marghoob, A. Rossi, A. C. Halpern, and K. S. Nehal, "Reflectance confocal microscopy of skin in vivo: From bench to bedside," Lasers in Surg. Med. **49**(1), 7–19 (2017).
4. 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).
5. I. Alarcon, C. Carrera, J. Palou, L. Alos, J. Malvehy, and S. Puig, "Impact of in vivo reflectance confocal microscopy on the number needed to treat melanoma in doubtful lesions," Br. J. Dermatol. **170**, 802–808 (2014).
6. I. Alarcon, C. Carrera, L. Alos, J. Palou, J. Malvehy, and S. Puig, "In vivo reflectance confocal microscopy to monitor the response of lentigo maligna to imiquimod," J. Am. Acd. Dermatol. **71**, 49–55 (2014).
7. P. Guitera, L. E. Haydu, S. W. Menzies, R. A. Scolyer, A. Hong, G. B. Fogarty, F. Gallardo, and S. Segura, "Surveillance for treatment failure of lentigo maligna with dermoscopy and in vivo confocal microscopy: new descriptors," Br. J. Dermatol. **170**(6), 1305–1312(2014).
8. J. Champin, J. L. Perrot, E. Cinotti, B. Labeille, C. Douchet, G. Parrau, F. Cambazard, P. Seguin, and T. Alix, "In vivo reflectance confocal microscopy to optimize the spaghetti technique for defining surgical margins of lentigo maligna," Dermatol. Surg. **40**(3), 247–256 (2014).
9. B. P. Hibler, M. Cordova, R. J. Wong, and A. M. Rossi, "Intraoperative real-time reflectance confocal microscopy for guiding surgical margins of lentigo maligna melanoma," Dermatol. Surg. **41**, 980–983 (2015).

10. B. L. Luck, K. D. Carlson, A. C. Bovik, and R. R. Richards-Kortum, "An image model and segmentation algorithm for reflectance confocal images of in vivo cervical tissue," IEEE Trans. Image Processing **14**(9), 1265–1276 (2005).
11. S. Kurugol, J. G Dy, M. Rajadhyaksha, K. W. Gossage, J. Weissmann, and D. H. Brooks, "Semi-automated algorithm for localization of dermal/epidermal junction in reflectance confocal microscopy images of human skin," in SPIE BiOS. International Society for Optics and Photonics, 2011, 79041A.
12. S. Kurugol, M. Rajadhyaksha, J. G. Dy, and D. H. Brooks, "Validation study of automated dermal/epidermal junction localization algorithm in reflectance confocal microscopy images of skin," in SPIE BiOS. International Society for Optics and Photonics, 2012, 820702.
13. 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," Journal of the European Academy of Dermatology and Venereology **25**(5), 554–558 (2011).
14. E. Somoza, G. O. Cula, C. Correa, and J. B. Hirsch, Automatic Localization of Skin Layers in Reflectance Confocal Microscopy, Springer International Publishing, Cham, 141–150 (2014).
15. S. C. Hames, M. Ardigo, H. P. Soyer, A. P. Bradley, and T. W. Prow, "Anatomical skin segmentation in reflectance confocal microscopy with weak labels," in Digital image computing: techniques and applications (dICTA'2015), international conference on. IEEE, 2015, 1–8.
16. S. C. Hames, M. Ardigò, H. P. Soyer, A. P. Bradley, and T. W. Prow, "Automated segmentation of skin strata in reflectance confocal microscopy depth stacks," PloS one, **11**(4), e0153208 (2016).
17. K. Kose, C. Alessi-Fox, M. Gill, J. G. Dy, D. H. Brooks, and M. Rajadhyaksha, "A machine learning method for identifying morphological patterns in reflectance confocal microscopy mosaics of melanocytic skin lesions in-vivo," in SPIE BiOS. International Society for Optics and Photonics, 2016, 968908.
18. M. A. Harris, A. N. Van, B. H. Malik, J. M. Jabbour, and K. C. Maitland, "A pulse coupled neural network segmentation algorithm for reflectance confocal images of epithelial tissue," PloS one **10**(3), e0122368 (2015).
19. S. Ghanta, M. I. Jordan, K. Kose, D. H. Brooks, M. Rajadhyaksha, and J. G. Dy, "A marked poisson process driven latent shape model for 3d segmentation of reflectance confocal microscopy image stacks of human skin," IEEE Trans. Image Processing **26**(1), 172–184 (2017).
20. A. P. Raphael, T. A. Kelf, E. Wurm, A. V. Zvyagin, H. P. Soyer, and T. W. Prow, "Computational characterization of reflectance confocal microscopy features reveals potential for automated photoageing assessment," Experimental Dermatology **22**(7), 458–463 (2013).
21. E. Richtig, R. Hofmann-Wellenhof, D. Kopera, L. El-Shabrawi-Caelen, and V. Ahlgrimm-Siess, "In vivo analysis of solar lentigines by reflectance confocal microscopy before and after q-switched ruby laser treatment," Acta Dermato Venereologica, **91**(2), 164–168 (2011).
22. A. Dima, M. Scholz, and K. Obermayer, "Automatic segmentation and skeletonization of neurons from confocal microscopy images based on the 3-D wavelet transform," IEEE Trans. Image Process **11**, 790–801 (2002).
23. S. V. Patwardhan, A. P. Dhawan, and P. A. Relue, "Classification of melanoma using tree structured wavelet transforms," Computer Methods and Programs in Biomedicine **72**(3), 223–239 (2003).
24. C. Q. Yuan, Z. Peng, and X. P. Yan, "Surface characterisation using wavelet theory and confocal laser scanning microscopy," J. Tribol. **127**, 394–404 (2005).
25. Y. Hu, A. Shen, T. Jiang, Y. Ai, and J. Hu, "Classification of normal and malignant human gastric mucosa tissue with confocal raman microspectroscopy and wavelet analysis," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **69**(2), 378–382 (2008).
26. P. Moulin and J. Liu, "Analysis of multiresolution image denoising shemes using generalized Gaussian and complexity priors," IEEE Trans. Inform. Theory. **45**, 909–919 (1999).
27. G. V. Wouwer, P. Scheunders, and D. V. Dyck, "Statistical texture characterization from discrete wavelet representation," IEEE Trans. Image Proc. **8**, 592–598 (1999).
28. M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," IEEE Trans. Image Processing **11**(2), 146–158 (2002).
29. S. Yu, A. Zhang, and H. Li, "A review of estimating the shape parameter of generalized Gaussian distribution," J. Comput. Inf. Syst. **8**(21), 9055–9064 (2012).
30. K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," IEEE Trans. Circ. Syst., Video Techn. **5**, 52–56 (1995).
31. A. Halimi, H. Batatia, J L.Digabel, G. Josse, and J.-Y. Tourneret, "Technical report associated with the paper "Statistical modeling of reflectance confocal microscopy images and characterization of skin lentigo," Tech. Rep., University of Toulouse, France, Feb 2017.
32. S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1993).
33. N. A. C. Cressie and H. J. Whitford, "How to use the two sample t-test," Biometrical Journal **28**(2), 131–148 (1986).
34. C. Wendorf, "Manuals for univariate and multivariate statistics," Stevens Point, WI: University of Wisconsin, 2004.
35. R. Rakotomalala, "Comparaison de populations: Tests paramétriques," Bartlett test, 27–29 (2013).
36. V. E. Johnson, "Revised standards for statistical evidence," Proceedings of the National Academy of Sciences **110**(48), 19313–19317 (2013).
37. M. Wiltgen, A. Gerger, C. Wagner, and J. Smolle, "Automatic identification of diagnostic significant regions in confocal laser scanning microscopy of melanocytic skin tumors," Methods Inf. Med. **47**(1), 14–25 (2008).

1. Introduction

Reflectance confocal microscopy (RCM) is a non-invasive imaging technique that enables in-vivo visualization of the epidermis down to the papillary dermis in real time [1,2]. The development of this technology has taken about two decades and is currently granted the clinical status in various countries [3]. Early investigations focused on the identification of cell populations in different skin layers. Research on RCM images addressed three aspects: i) clinical studies to evaluate their usefulness, ii) segmentation of nuclei, and iii) classification of skin tissues. Later, interest has moved to diagnosis applications (especially cancer) with the objective of improving the sensitivity and specificity [4,5]. Other applications of RCM have been reported such as treatment follow up [6], surveillance of lentigo malign treatment [7,8], and guidance of cutaneous surgery [9].

At the time of this writing, a review of the literature on RCM imaging during the previous semester shows about 1500 publications. Most of them report methods based on visual inspection or ad'hoc quantitative measures. Research works reporting automatic techniques to process RCM images are limited. Luck et al. [10] have been pioneers in developing automatic RCM image processing. Their nuclei segmentation method was based on Gaussian image models for the nuclei reflectivity and the cytoplasm fibers intensity, and a Gaussian Markov random field for spatial correlation. They proposed a Bayesian classification algorithm to label tissues. Later, Kurugol et al. developed a semi-automatic method to locate the dermoepidermal junction (DEJ) using a statistical classifier of texture features describing the brightness of basal cells [11,12]. Other methods were then investigated for RCM image proc essing. Koller et al. [13] proposed a wavelet-based decision tree classification method to distinguish benign and malignant melanocytic skin tumors in RCM. The authors of [14] developed an automatic method to localize skin layers in RCM images based on texture analysis. Hames et al. developed a logistic regression classifier to automatically segment the different layers of the skin in RCM images [15,16]. In [17], SURF texture features were classified with support vector machines (SVM) to identify skin morphological patterns in RCM images. In [18], an algorithm based on neural networks was developed to segment nuclei in RCM images. All these works use machine learning techniques to classify texture features directly extracted from the image. More recently, a Bayesian method to segment the dermoepidermal junction in 3D RCM images has also been proposed [19]. This method is based on a marked Poisson process to detect locations of papillae, a spherical model for their shape and a Gaussian model of texture features to capture their appearance. Determining quantitative markers for tissue characterization in RCM images has not been widely investigated. Raphael et al. [20] reported a characterization method of RCM images to assess photoageing. In their study, features including the intensity, 2D wavelet coefficient values, 2D Fourier coefficients and shapes were correlated with clinical data. They concluded that the image intensity and the wavelet coefficients have no significant correlation, contrary to Fourier coefficients and shapes. In [21], Richtig et al. reported that solar lentigines show rete-edges with regular honeycomb patterns and edged dermal papillae in RCM images. Based on these observations, we propose in this paper to characterize solar lentigo using a parametric statistical model for textures associated with RCM images. More precisely, we show that a quantitative parameter of this model allows identifying lentigo in these images.

The proposed characterization method is based on a multi-resolution discrete wavelet decomposition of RCM images. This technique has been considered for tissue characterization and classification in many practical applications using directly values of wavelet coefficients [22–25]. However, wavelet coefficients have a large variability due to image details such as the range of intensities, resolution and contrast. In addition, RCM images have a high resolution ($50 \times 1000 \times 1000$) leading to processings with high computational complexity. The proposed method fits within the framework of machine learning. The aim is to address the two aforementioned limitations. A generalized Gaussian distribution (GGD) model is proposed to represent wavelet coefficients at each depth of the skin. This model maps the data into a two-parameter space resulting in a strong

dimension reduction and allowing the implementation of a real-time lentigo detection algorithm. Precisely, an SVM classifier is trained offline for each depth. It allows detecting in real-time the depth at which the lentigo is present. In addition, the multiresolution approach has the advantage of alleviating the problem of dependency to the resolution of the image. Experiments with real data from a clinical study show that the shape parameter of the GGD (alone) is a good indicator to discriminate between lentigo and healthy tissues. These experiments show that lentigo is identifiable at depths between $50\mu m$ and $60\mu m$. The remainder of the paper is organized as follows. Section 2 presents the proposed method for lentigo characterization and classification. It describes the wavelet decomposition process and the statistical model used for the wavelet coefficients. It also presents the proposed texture characterization method based on T-tests and the Bayes factor and the SVM classification protocol. Experimental results are presented in Section 3. Discussions reported in Section 4 summarize the findings of this study with some physical interpretation. Conclusions and perspectives for future work are finally reported in Section 5.

2. Proposed method

This section presents a new statistical method to detect lentigo in RCM and simultaneously identify the depth where the lesion is best identifiable. The method has very low computational cost and is suitable for real-time detection. Let $\boldsymbol{Y}_k \in \mathbb{R}^{P\times N\times N}$ denotes a stack of P images of $N \times N$ pixels from a patient k. The tensor Y_k contains P images denoted as $Y_{k,1}, \ldots, Y_{k,P}$, where $Y_{k,d}$ is a grayscale image, oriented in the horizontal (en face) plane, associated with the skin depth d. Image registration ensures that for all patients, images acquired at depth d correspond to the same measure of depth in micrometers. This paper considers the problem of detecting a lentigo in an image stack Y_k in real-time, while identifying the depth at which the lesion is best identifiable. Formally, we define a label z_k associated with the patient k that takes the value 1 when the stack of images Y_k has a lentigo and 0 otherwise. Let $\{\boldsymbol{Y}_k^{(0)}\}_{k=1}^{K_0}$ and $\{\boldsymbol{Y}_k^{(1)}\}_{k=1}^{K_1}$ denote some expert annotated stacks of images corresponding to healthy and lentigo skins. We first consider the identification of the C most characteristic depths ($1 \leq C \leq P$). We consider then the classification problem consisting of estimating the label of a collection of C test images $\left\{Y_{k,d_1}, \ldots, Y_{k,d_C}\right\}$, at a given range of depths $[d_1, d_C] \subset [1, P]$, denoted as z_k using a training set composed of the annotated characteristic depth images. The proposed method is organized in a pipeline depicted in Fig. 1, whose components are detailed below.

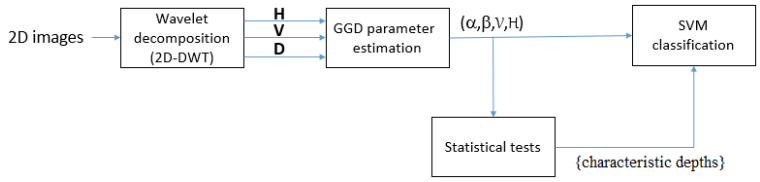

Fig. 1. Block-diagram of the proposed method. It consists of 4 different stages. First, a wavelet decomposition is applied to the image. Second, GGD parameters (α, β) are estimated, along with the variance V and the entropy H. Third, statistical tests are implemented to identify characteristic depths. Finally, SVM classifiers based on the GGD parameters classify tissues as lentigo or healthy.

2.1. Wavelet decomposition

Inspecting visually the considered images, one can easily notice that the lentigo produces a significant change in the texture of the images (Fig. 2). At the DEJ, lentigines exhibit more papillae whose shapes appear more irregular than on healthy skin, explaining this difference

in texture. As mentioned earlier, wavelet coefficients have proven effective in capturing texture properties in various applications [22–25]. Following this approach, the first stage of our method consists in applying a wavelet decomposition to RCM images.

A Daubechies wavelet filter bank is used to decompose each image Y_d. Such filter bank is computationally efficient and can be implemented in real-time. The decomposition is performed at four scales. At each scale, the result is a set of horizontal, vertical and diagonal coefficients that we denote by $\boldsymbol{H}$, $\boldsymbol{V}$ and $\boldsymbol{D}$, respectively. Let $\mathbf{x}_s^{(b)}$ denotes the $L \times 1$ vector of the arbitrarily ordered coefficients obtained at the scale s for the band $b \in \{\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{D}\}$. Figure 3 shows an example of decomposition of an RCM image at scale 1. In this work, we are interested in identifying the lentigo lesion in real-time based on these coefficients. Given the large number of data, the next section addresses the problem of dimension reduction.

2.2. Statistical modeling for dimension reduction

Designing a classification method directly from the wavelet coefficients would be ineffective due their variability (see [20]) and inefficient due the computational cost of the required learning algorithm. It is therefore necessary to map the wavelet coefficients data into a lower dimensional space. Such representation should uncover the hidden structure that discriminates healthy and lentigo images. It should also allow the implementation of a simple classification method for rapid detection. In this work, as in [26–29], we propose to represent the empirical distribution of the coefficients $\mathbf{x}_s^{(b)}, \forall b \in \{\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{D}\}, \forall s \in \{1, \ldots, 4\}$ using a zero-mean generalized Gaussian distribution (GGD). This distribution is known to capture the statistical properties of wavelets for a large class of images [28, 30]. It allows us to reduce the size of the features vector from several thousand coefficients to two parameters only, while capturing the variability of the coefficients. The GGD has the following probability density function

$$f(x; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left|\frac{x}{\alpha}\right|^{\beta}\right) \tag{1}$$

where $\alpha \in \mathbb{R}^+$ is a scale parameter, $\beta \in \mathbb{R}^+$ is a shape parameter that controls the density tail, and $\Gamma(\cdot)$ is the gamma function. We estimate the values of α and β for each set $\mathbf{X}_s^{(b)}, \forall b \in \{\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{D}\}, \forall s \in \{1, ..., 4\}$ by maximum likelihood estimation using a Newton-Raphson algorithm [28].

Let $\boldsymbol{\delta}_{b,s}^{(d)} = \left[\alpha_{b,s}^{(d)}, \beta_{b,s}^{(d)}\right]$ be the (1×2) parameter vector estimated for depth d at scale s from band b. For depth d and scale s, the image is represented by a parameter vector $\theta_s^{(d)} = \left\{\boldsymbol{\delta}_{s,\boldsymbol{H}}^{(d)}, \boldsymbol{\delta}_{s,\boldsymbol{V}}^{(d)}, \boldsymbol{\delta}_{s,\boldsymbol{D}}^{(d)}\right\}$ with s=1,...,4. We will denote as $\boldsymbol{\theta}^{(d)} = \left\{\theta_1^{(d)}, \ldots, \theta_4^{(d)}\right\}$ the parameter vector associated with the depth d. This proposed dimension reduction step results in representing a patient by an array of parameters $\boldsymbol{\Theta} = \left\{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(P)}\right\}$, giving a total number of parameters of only $2 \times 3 \times 4 \times P$. In addition, we will experimentally show that i) only few depths ($C \ll P$) will be necessary to identify the lentigo; ii) one single parameter out of the two (α, β) is sufficient to classify healthy and lentigo patients; iii) scale 1 has enough information for lentigo identification; and iv) the 3 bands are better considered jointly. This results in a more reduced dimension, allowing a patient to be represented by only $3 \times C$ parameters. We will also show in Section 3.2 that the density (1) fits correctly (with good Kolmogorov-Smirnov test) the empirical distribution of the wavelet coefficients. The next section investigates a statistical test using the estimated parameters (α, β) to assess their ability to discriminate healthy from lentigo patients.

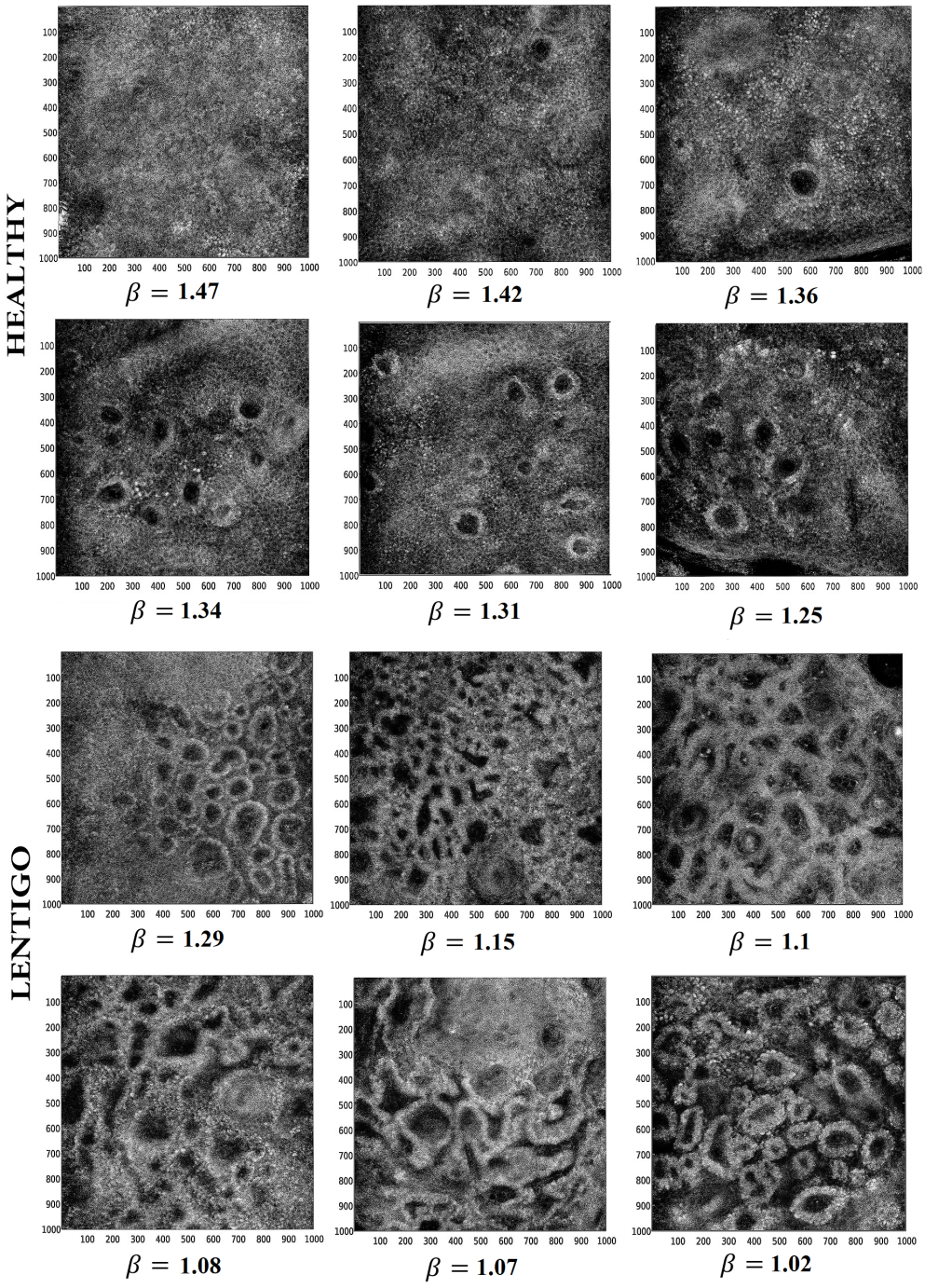

Fig. 2. Typical images at the DEJ from six healthy (left to right and top to bottom #1, #5, #4, #3, #2, #6) and six lentigo (#33, #38, #40, #31, #44, #37) patients. One can observe coarse texture in the form of round shapes in the presence of lentigo. N.B. Values of the parameter β are explained in section 4.

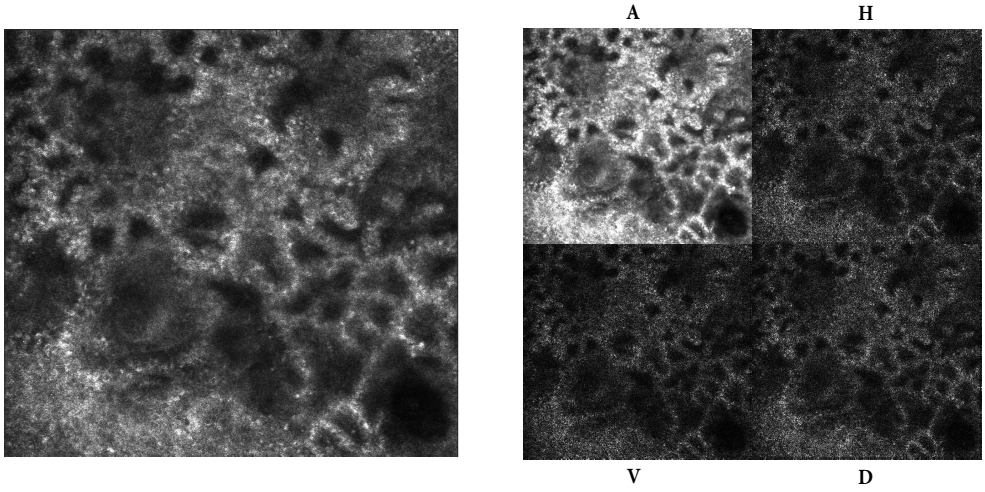

Fig. 3. Example of the first scale wavelet decomposition (right) of an RCM image (left). The decomposition has four bands: Approximation ($\boldsymbol{A}$), Horizontal ($\boldsymbol{H}$), Vertical ($\boldsymbol{V}$), Diagonal ($\boldsymbol{D}$). Applying the same scheme to the approximation gives the next decomposition scale. As explained in the text, our statistical method consists of estimating the GGD parameters (α, β) for bands $(\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{D})$ at each scale. These parameters are used for the characterization and classification of the underlying tissues.

2.3. Lentigo characterization by parametric T-tests

Before applying a classification method, we performed a parametric T-test to the shape and scale parameters α and β of the different patients. This test allows assessing the statistical significance of these parameters to separate healthy and lentigo patients. Note that similar tests were also performed with the entropy and the variance of the distribution. The corresponding results are available in a technical report [31] and are not presented here for brevity. We consider a corpus composed of images from n healthy and m lentigo patients, annotated by a dermatologist. Let $\boldsymbol{\Theta}^{(\mathrm{S})} = \left\{\boldsymbol{\Theta}_1^{(\mathrm{S})}, \ldots, \boldsymbol{\Theta}_n^{(\mathrm{S})}\right\}$ and $\boldsymbol{\Theta}^{(\mathrm{L})} = \left\{\boldsymbol{\Theta}_1^{(\mathrm{L})}, \ldots, \boldsymbol{\Theta}_m^{(\mathrm{L})}\right\}$ denote the parameters estimated for healthy and lentigo patients, respectively. The maximum likelihood estimator (MLE) is known to be asymptotically Gaussian and asymptotically efficient [32, p.191]. Thus for large sample size, these parameters can be assumed to be distributed according to Gaussian distributions. Figure 4 shows that this assumption is reasonable for all the parameters. As an illustration, we consider here the T-test for the parameter α at an arbitrary depth. Let $\mu_{b,s}^{(\mathrm{S})}$ denote the mean of the parameters $\left\{\alpha_{b,s}^1, \ldots, \alpha_{b,s}^n\right\}$ estimated for the healthy patients for the band b at the scale s. Similarly, $\mu_{b,s}^{(\mathrm{L})}$ is the mean of the parameters $\left\{\alpha_{b,s}^1, \ldots, \alpha_{b,s}^m\right\}$ estimated for lentigo patients at the same depth and scale. A classic two-sample T-test [33–35] has been used to compare the means $\mu_{b,s}^{(\mathrm{S})}$ and $\mu_{b,s}^{(\mathrm{L})}$

$$H_{b,s}^0 : \mu_{b,s}^{(\mathrm{S})} = \mu_{b,s}^{(\mathrm{L})}, \tag{2}$$

$$H_{b,s}^1 : \mu_{b,s}^{(\mathrm{S})} \neq \mu_{b,s}^{(\mathrm{L})}. \tag{3}$$

In this study, we chose a probability of false alarm $PFA = 0.05$ corresponding to a threshold $T_{PFA} = 2.02$ to reject $H_{b,s}^0$. The p-value has also been calculated for each test. Then, the following

decision rules have been applied

- When p value > 0.10 $\longrightarrow$ the observed difference is "not significant"
- When p value $\in [0.05, 0.10]$ $\longrightarrow$ the observed difference is "marginally significant"
- When p value $\in [0.01, 0.05[$ $\longrightarrow$ the observed difference is "significant"
- When p value < 0.01 $\longrightarrow$ the observed difference is "highly significant"

Following [36], we also calculated the Bayes factor (BF) given by

$$BF = \left(\frac{\nu + T}{\nu + \left(T - \sqrt{(\nu\gamma^*)}\right)^2} \right)^{(n+m)/2} \tag{4}$$

where ν is the degrees of freedom, T the T-test score, $\gamma^* = \gamma^{2/(n+m-1)} - 1$ and $\gamma = [(T^2_{PFA}/\nu) + 1]^{(n+m)/2}$. With this statistics the hypothesis $H^0_{b,s}$ is rejected when $BF > \sqrt{\nu\gamma^*}$. In the experimentation section, we will show that β is the only parameter that can be used for detecting lentigo at specific depths, meaning that β is a good tissue discriminator for those depths.

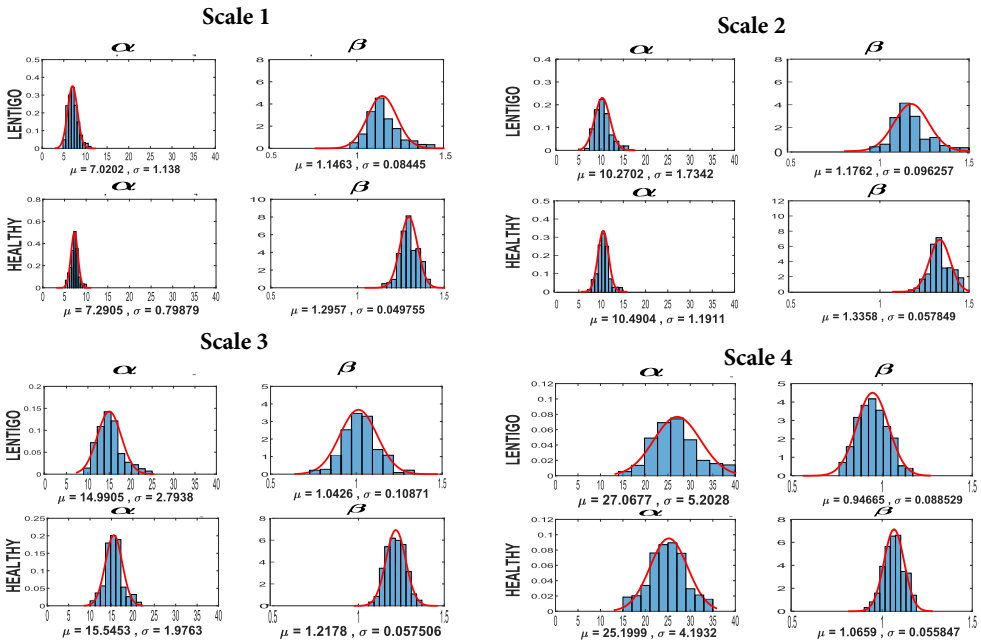

Fig. 4. Histograms of the scale and shape parameters, estimated from band $\boldsymbol{H}$ at depth $54\mu m$, with their means and standard deviations for all healthy and lentigo patients. Similar histograms are obtained for other bands and scales.

2.4. *Lentigo detection by SVM classification*

The last stage of the proposed method consists in classifying patients as lentigo or healthy based on the GGD parameters. Precisely, the GGD parameters, the variance, and entropy associated with the C most characteristic depths $\boldsymbol{d} = (d_1, \ldots, d_C)$, as identified by the statistical test (presented in Section 2.3), are considered. In our experiments, we chose $C = 3$ and $\boldsymbol{d} = (d_1, d_2, d_3)$. Many

classification methods could be investigated to classify lentigo and healthy images. We present here the results obtained with a linear SVM classifier that is simple to implement and has shown good performance in many applications. Our objective here is to confirm that the parameter β is able to classify tissues better than α, the variance (V), and the entropy (H). Therefore, we have considered these parameters separately and have determined the corresponding classification performance. We also aim at identifying which scale provides better classification and whether there exists any preferred direction detail, horizontal, vertical or diagonal. The parameters were estimated for each scale $s \in \{1, 2, 3, 4\}$ from each band $b \in \{\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{D}\}$. Let us consider the case of β as an example. The estimated parameters β for each scale and band are summarized in the following matrix

$$\begin{bmatrix} \beta_{\boldsymbol{H},1} & \beta_{\boldsymbol{H},2} & \beta_{\boldsymbol{H},3} & \beta_{\boldsymbol{H},4} \\ \beta_{\boldsymbol{V},1} & \beta_{\boldsymbol{V},2} & \beta_{\boldsymbol{V},3} & \beta_{\boldsymbol{V},4} \\ \beta_{\boldsymbol{D},1} & \beta_{\boldsymbol{D},2} & \beta_{\boldsymbol{D},3} & \beta_{\boldsymbol{D},4} \end{bmatrix} \tag{5}$$

where $\boldsymbol{\beta}_{b,s} = \left(\beta_{b,s}^{(d_1)}, \ldots, \beta_{b,s}^{(d_C)}\right)$ is the vector of parameters estimated from the C images, for scale s and band b. SVM classifiers were designed for each element of the matrix (5), for each row, each column and for the whole matrix (5) leading to a total of 20 classifiers. In other words, the scales were considered separately and jointly, and for each scale the bands were also considered separately and jointly. The leave-one-out method was used to evaluate the classification performance. This operation was conducted 1000 times, and all results were averaged in order to obtain averaged confusion matrix and performance indicators (sensitivity, specificity, precision and accuracy). In the experiments presented in Section 3.5, we will focus on the 20 SVM classifiers that are based on the parameter β (results obtained with the other parameters are available in [31]).

3. Experiments

3.1. Data

RCM imaging was performed using an apparatus Vivascope 1500. The in-vivo images were acquired from the stratum corneum, the epidermis layer, the dermis-epidermis junction (DEJ) and the upper papillary dermis. Each RCM image shows a $500 \times 500 \mu m$ field of view (1000×1000 pixels). Forty-five women aged 60 years and over were recruited. The study was conducted according to the principles of the declaration of Helsinki, and was approved by the *Comité de Protection des Personnes Sud-Ouest et Outre-mer* III in Bordeaux, France, No. CPP 2011/36. All the volunteers gave their informed consent for examination of skin by RCM. According to the clinical evaluation performed by a physician, volunteers were divided into two groups. Twenty-seven ($m = 27$) women with at least 3 lentigines on the back of the hand formed the first group and eighteen ($n = 18$) women with no lentigo formed the control group. For each volunteer, the acquisition on the same location was repeated twice, giving two stacks for each patient. Images were taken on lentigo lesions for volunteers of the first group and on healthy skin on the back of the hand for the control group. An examination of each acquisition was performed in order to locate the stratum corneum and the DEJ precisely on each image. Consequently, our database contained $M = 45$ patients. For each patient, we retained two stacks of 25 RCM images giving a total of 2250 images. Images were registered in order to correspond to the same depths.

3.2. Statistical analysis of wavelet coefficients

This section illustrates the goodness of fit of the generalized Gaussian distribution for the wavelet coefficients of RCM images. Figure 5 compares the histograms of the wavelet coefficients from band $\boldsymbol{H}$ with the estimated GGD distributions for the four scales at 3 representative depths. This figure concerns two arbitrary healthy and lentigo patients, namely patients #6 and #38

respectively. It shows the good fit between the observed histograms and the estimated distributions for all scales for both healthy and lentigo images. For scale 4, the distribution misses the mode but fits well the shape. This can be explained by the small number of data since coefficients at this scale are very sparse and tend toward a flat histogram. Slight differences in the shape of the distributions can be observed between healthy and lentigo patients, as illustrated by the differences in the corresponding parameters α and β. These differences are at the basis of the proposed characterization method as illustrated in the next experiments. Good results have also been obtained for the bands $\boldsymbol{V}$ and $\boldsymbol{D}$. They are not shown here for brevity.

Figure 6 shows the quantitative assessment of the fit using the the Kolmogorov-Smirnov (KS) test. The mean KS statistic score of the whole population (45 patients) has been calculated at each depth for all scales. One notices the excellent scores with values of the KS statistic very close to zero. Quantitatively, considering all depths and bands, the mean KS is 0.008 ± 0.002 for scale #1; 0.009 ± 0.003 for scale #2; 0.013 ± 0.004 for scale #3; and 0.027 ± 0.009 for scale #4. The increase of the score with the scale is due to sparser data.

3.3. Identification of characteristic depths

Each of the 2250 images was decomposed according to the multi-resolution analysis presented in Section 2.1 and GGD distributions were fitted to each scale at each band. Having acquired two stacks of 25 images for each patient, one of the two stacks was selected randomly for the analysis. The mean $\mu_{b,s}^{(\mathrm{S})}$ and $\mu_{b,s}^{(\mathrm{L})}$ for α and β were calculated for healthy and lentigo images, respectively. The curves showing the evolution of the average parameters with respect to the image depth were finally elaborated. To account for variability, the process of selecting one stack for each patient was repeated 300 times and average curves with standard deviations were calculated. Results are shown in Fig. 7, which clearly shows that, for all depths, bands and scales, α is not characteristic. Conversely, β allows the discrimination of healthy and lentigo images for all bands at scales 1 to 3 for depths between $31\mu m$ and $76\mu m$. Depths around $50\mu m$ give the largest discrimination. Results obtained with the bands $\boldsymbol{H}$ and $\boldsymbol{V}$ for the fourth scale show large variability due to the low quantity of data. Figure 2 shows two sets of images associated with six healthy (#1, #2, #3, #4, #5, #6) and six lentigo (#31, #33, #37, #38, #40, #44) patients. As expected, one notices that in the presence of lentigo images are more textured at the DEJ range of depths.

3.4. Statistical significance with T tests

The parametric test described in Section 2.3 has been applied to the parameters of each band, scale and depth to assess the significance of the results. Figures 8 and 9 show the p-values and Bayes factors associated with the two-sample t-tests conducted respectively with α and β, at different depths. The p-value has been represented in $-\log$ scale for readability. Figure 8 shows weak values of both p-value and Bayes factor confirming that α cannot discriminate healthy and lentigo images. Figure 9 shows high scores for both indicators for a range of depths. Table 1 presents depths that give T_β higher than the threshold, hence confirming the hypothesis that the parameter β can be used to discriminate healthy and lentigo patients. This table also shows the depths that provide p-values lower than the probability of false alarm $PFA = 0.05$ and their corresponding Bayes factor. It can be seen that for all depths where $H_{\cdot,\cdot}^0(\beta)$ is rejected, p-values are lower than 0.01. According to our decision rules, the results are *highly significant* for depths $49\mu m$ and $60\mu m$ for scale one, with the highest score at $54\mu m$. These results are in good agreement with the quantitative differences shown in Fig. 7. These results confirm that β gives a good test statistics for discriminating lentigo and healthy skin, especially at depths around $50\mu m$. As mentioned in the introduction, lentigines are mainly characterized in RCM by the disorganization of the dermoepidermal junction (DEJ). It is therefore not surprising to find no significant difference between healthy and lentigo skin in the stratum corneum and higher epidermis layers (below $40\mu m$). Coherently, the parameter β is very discriminant at depths close

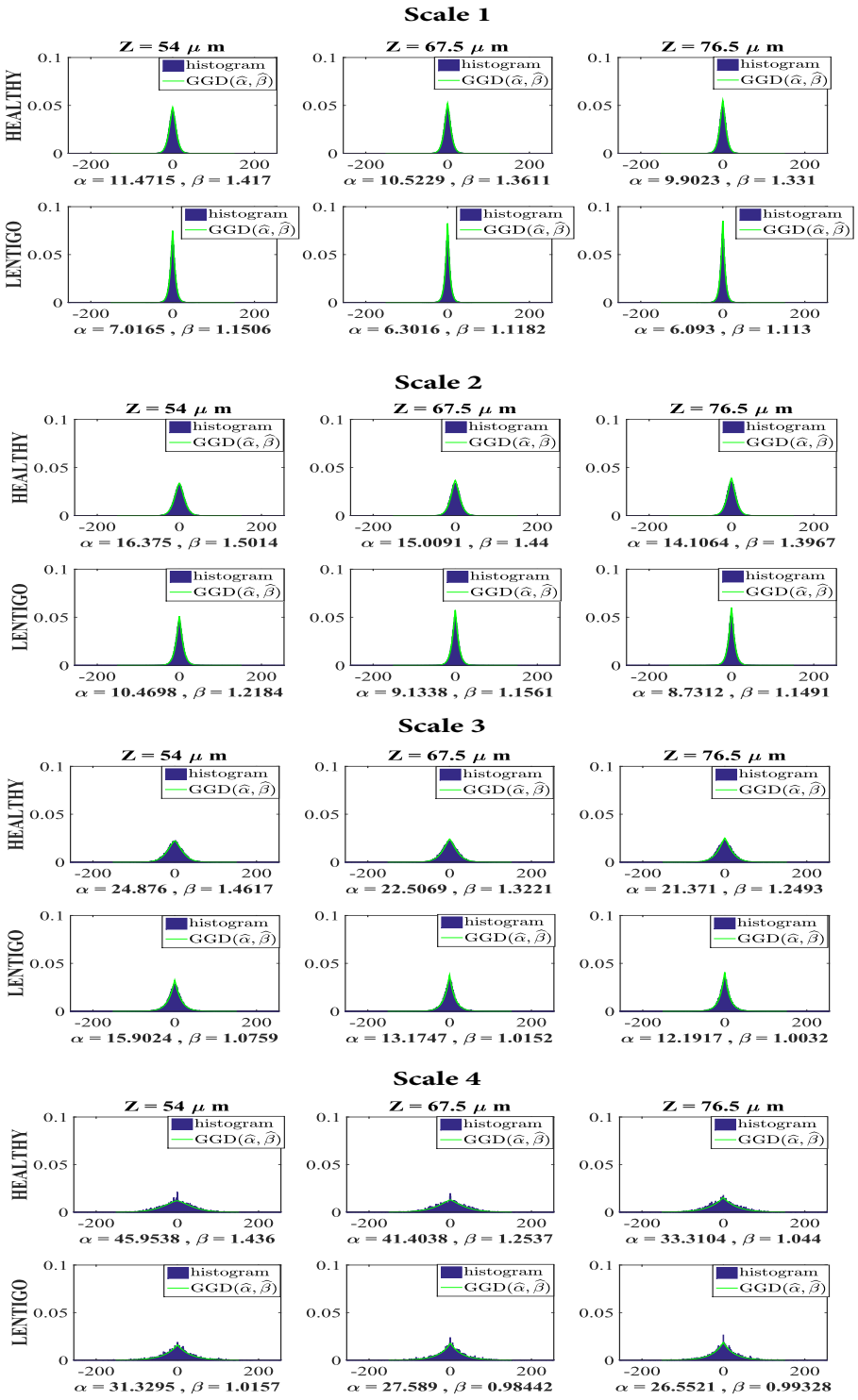

Fig. 5. Histograms of the wavelet coefficients from band $\boldsymbol{H}$ at the four scales and the corresponding estimated GGD distributions. The figure shows data from two arbitrary healthy and lentigo patients (#6 and #38 respectively) at three representative depths (one depth per column).

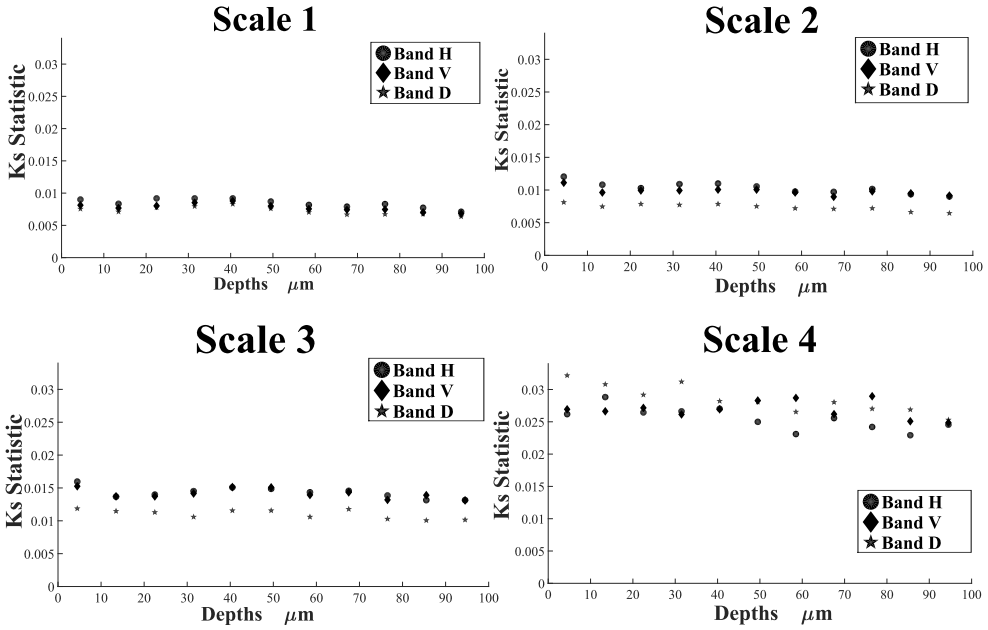

Fig. 6. Assessment of the GGD fit to wavelet coefficients. Mean KS statistic for the whole population at some selected depths, shown by scale and band. Scores are very good for all configurations, although they increase with higher scales, due to sparser data.

to $50\mu m$, which corresponds to the average location of the DEJ as shown in Fig. 10.

3.5. Performance of the SVM classification

For the classification, we processed separately three images ($C = 3$) corresponding to characteristic depths ($49.5\mu m, 54\mu m, 58.5\mu m$) for each patient. As described in section 2.4, a GGD distribution was fitted at each scale (#1, #2, #3, #4) to wavelet coefficients form separate bands ($\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{D}$). The 2 parameters ($\alpha, \beta$) were estimated, along with the variance and the entropy. For each of these parameters, SVM classifiers were trained and tested to classify patients into 2 classes referred to as "lentigo" and "healthy". The leave-one-out cross-validation method was used to compute the different probabilities of errors. This method uses $M - 1$ images for training (where $M = 45$ is the number of patients in the database) and the remaining image for testing. This operation was run $N = 1000$ times. For each experiment, we considered only images from one acquisition out of the available two stacks (for each patient). The obtained N results were used to calculate the average confusion matrix shown in Table 2 and to evaluate average indicators (sensitivity, specificity, precision and accuracy). This table allows us to assess the classification performance for each scale and for the three bands separately and jointly. The results show that the classification of healthy and lesion tissues is similar for the different combinations of detail-bands, and thus we recommend to use the joint bands ($\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{D}$). Regarding the different scales used for the analysis, Table 2 shows that results from the first scale are slightly better than those obtained with the other scales. Thus we recommend to use the first scale of the wavelet decomposition. Our method was then compared to the method presented in [13], named here Koller. This method consists of extracting 39 features from each RCM image (more technical details are available in [37]) and applying a classification procedure based on a regression tree (CART). We tested this method on the real RCM images using a leave one out procedure. As shown in Table 3, the accuracy (75.5%) obtained with the Koller method is smaller than the one obtained with the proposed method and

leads to three additional misclassified patients. In addition, this result is obtained at the expense of higher complexity given the number of parameters (39 for Koller as compared to our method).

For illustration purpose, Fig. 11 shows examples of classified RCM images using the proposed methodology. The results concerning the other parameters (α, Entropy, Variance) were not convincing. Thus, they are not reported in this paper.

Table 1. Depths where $H^0_{\cdot,\cdot,\cdot}(\beta)$ is rejected (t-score > $T_{0.05} = 2.02$); the corresponding p-value and Bayes factor (BF) are shown. The first row gives intervals of depths (min depth to max depth) where T-scores are significant. The second row shows depths giving highest T-scores (maximal T-score ∓ 10%). The third row shows the depths corresponding to the maximal T-score. P-values and Bayes factors corresponding to each depth are shown below.

		scale 1		scale 2		scale 3		scale 4	
		min	max	min	max	min	max	min	max
$T_\nu^{(\beta)} > 2.02$	depths (μm)	14	90	18	90	18	90	18	94
	T-score	2.10	2.22	2.13	2.27	2.10	2.20	2.14	2.11
	p-value	0.040	0.037	0.040	0.030	0.040	0.037	0.038	0.040
	BF	11.00	13.40	7.18	15.23	23.10	29.70	13.24	10.20
Highest $T_\nu^{(\beta)}$	depths (μm)	48	60	48	63	48	68	48	68
	T-score	4.80	4.84	4.54	4.73	4.02	4.07	3.99	4.26
	p-value	0.0002	0.0002	0.0004	0.0003	0.002	0.002	0.003	0.001
	BF	877	970	891	949	454	460	359	435
Maximal $T_\nu^{(\beta)}$	depths (μm)	54		54		59		59	
	T-score	5.11		4.94		4.69		4.41	
	p-value	0.00008		0.00016		0.00031		0.00074	
	BF	1246		1459		436		507	

Table 2. Confusion matrices of SVM classifiers based on the vector of parameters ($\beta_{49.5\mu m}, \beta_{54\mu m}, \beta_{58.5\mu m}$) from the three characteristic depths. Twenty different classifiers have been tested with the leave-one-out method for the $\boldsymbol{H}$, $\boldsymbol{V}$, $\boldsymbol{D}$, and $\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{D}$ bands at the four scales 1, 2, 3, 4 separately and combined (1, 2, 3, 4). Se and Sp stand for the sensitivity and specificity. One notices the good accuracy for all bands (there is no preferred direction, justifying the joint $\boldsymbol{H}, \boldsymbol{V}, \boldsymbol{D}$ band) especially with the first scale.

$\boldsymbol{\beta}$		Scale 1			Scale 2			Scale 3			Scale 4			All scales		
Confusion matrix		L	$\hat{h}$	Se Sp	$\hat{L}$	$\hat{h}$	Se Sp	$\hat{L}$	$\hat{h}$	Se Sp	$\hat{L}$	$\hat{h}$	Se Sp	$\hat{L}$	$\hat{h}$	Se Spf
H	Lentigo	22	5	81.4 %	21	6	77.7 %	20	7	74%	19	8	70.3 %	21	6	77.7 %
	Sain	3	15	83.3 %	4	14	77.7 %	4	14	77.7%	6	12	66.6 %	5	13	72.2 %
	Precision	88%	75%		84%	70%		83.3 %	66.6 %		76%	60%		80.7 %	68.4 %	
	Accuracy	82.2%			77.7%			75.5 %			68.8 %			75.5 %		
V	Lentigo	22	5	81.4%	20	7	74%	20	7	74%	20	7	74%	20	7	74%
	Sain	3	15	83.3%	4	14	77.7%	4	14	77.7%	5	13	72.2 %	5	13	72.2 %
	Precision	88%	75%		83.3 %	66.6 %		83.3%	66%		80%	65%		80%	65%	
	Accuracy	82.2%			75.5 %			75.5 %			73.3 %			73.3%		
D	Lentigo	22	5	81.4 %	21	6	77.7 %	20	7	74%	19	8	70.3 %	21	6	77.7 %
	Sain	3	15	83.3 %	5	13	72.2 %	5	13	72.2%	5	13	72.2%	5	13	72.2 %
	Precision	88 %	75 %		80.7 %	68.4 %		80%	65%		79.1 %	61.9 %		80.7 %	68.4 %	
	Accuracy	82.2 %			75.5 %			73.3 %			71.1 %			75.5 %		
HVD	Lentigo	22	5	81.4%	21	6	77.7%	20	7	74%	19	8	70.3%	21	6	77.7 %
	Sain	3	15	83.3%	4	14	77.7%	4	14	77.7%	5	13	72.2%	5	13	72.2 %
	Precision	88%	75%		84%	70%		83.3%	66%		79.1%	61.9%		80.7 %	68.4 %	
	Accuracy	82.2%			77.7%			75.5 %			71.1 %			75.5 %		

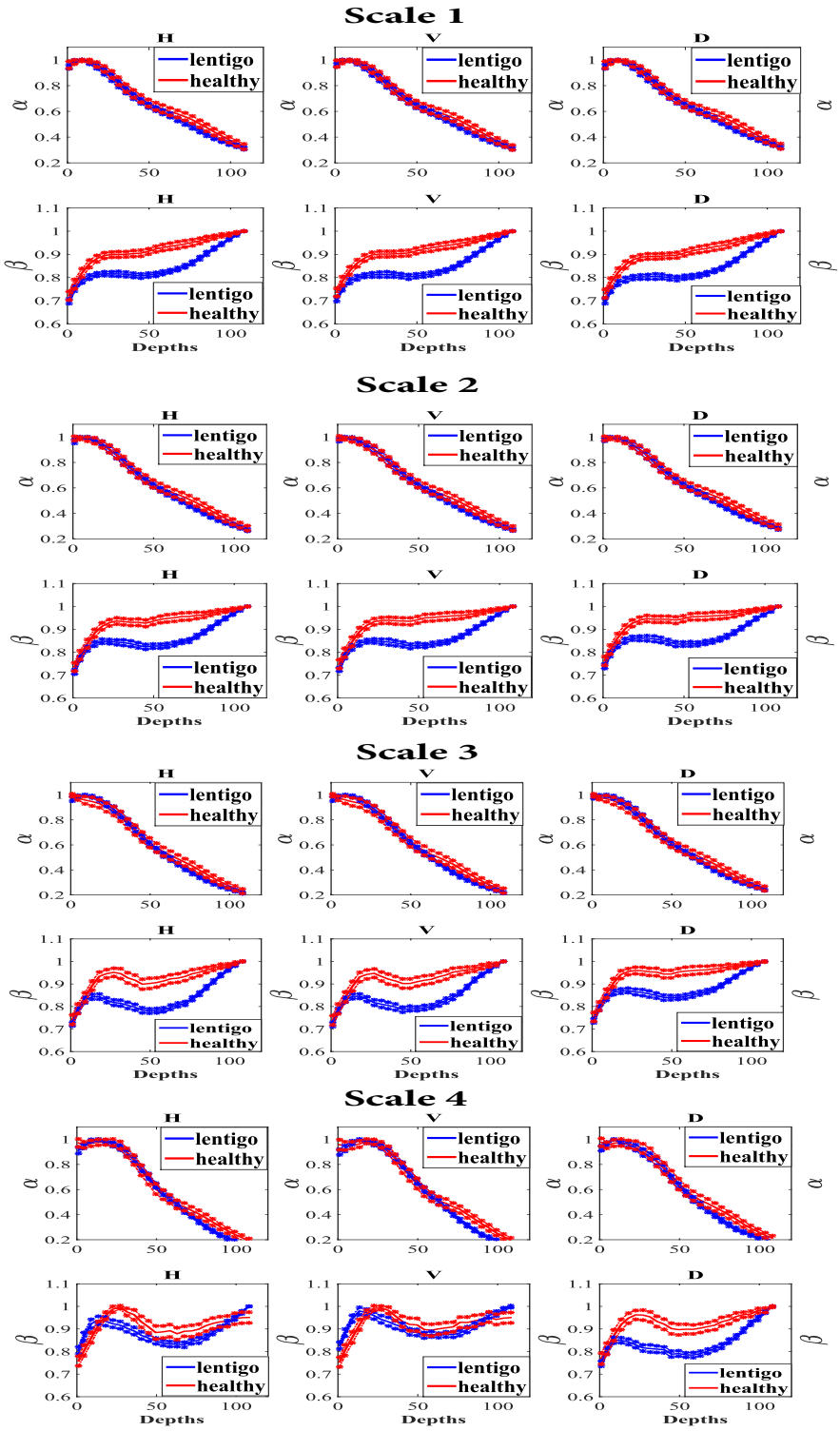

Fig. 7. Evolution of the average parameters $\hat{\alpha}$ and $\hat{\beta}$ throughout the depth for the different bands at all scales. Values of α are too similar for healthy and lentigo patients and cannot be used for discrimination. The parameter β shows significant difference for depths between $31\mu m$ and $76\mu m$, with maximal difference at around $50\mu m$. Our conclusion is that this parameter β can discriminate healthy and lentigo skin tissues.

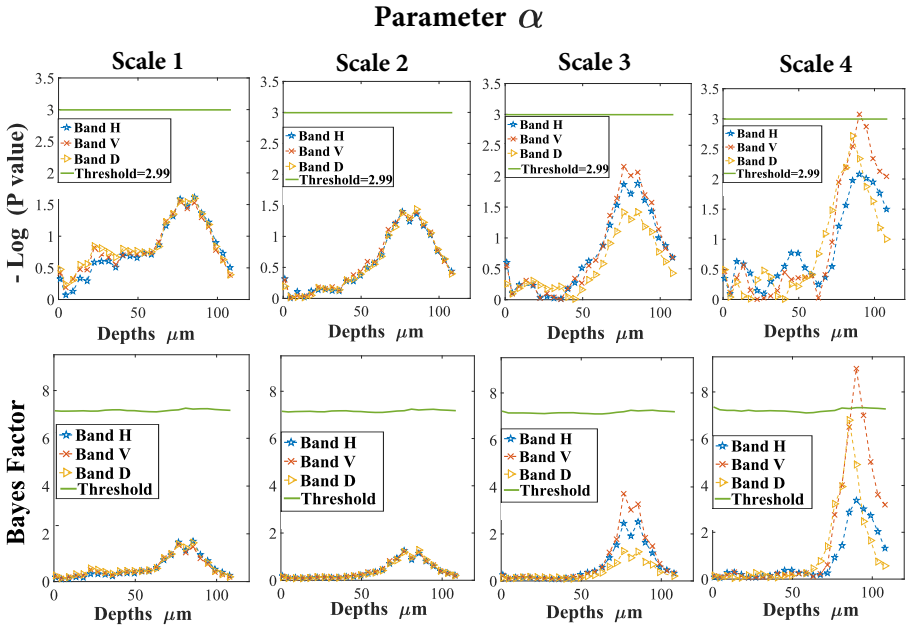

Fig. 8. P-value (in $-$ log scale) and Bayes factor of the T test for α. The weak scores show that α is clearly not a discriminant between healthy and lentigo images.

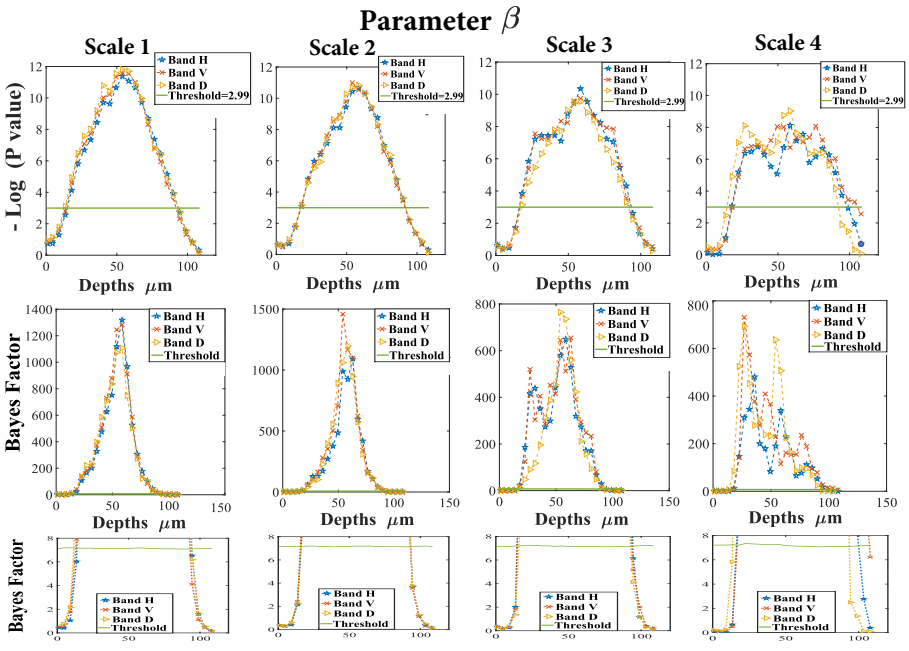

Fig. 9. P-value (in $-log$ scale) and Bayes factor (BF) of the T test for β. The third row contains zooms on the lower scores of the BF to clarify the significance threshold. Strong scores can be seen for depths between $31\mu m$ and $76\mu m$. Highest scores are obtained with depths around $50\mu m$. This confirms that β is a good discriminant function that can be used to separate healthy and lentigo images at these depths.

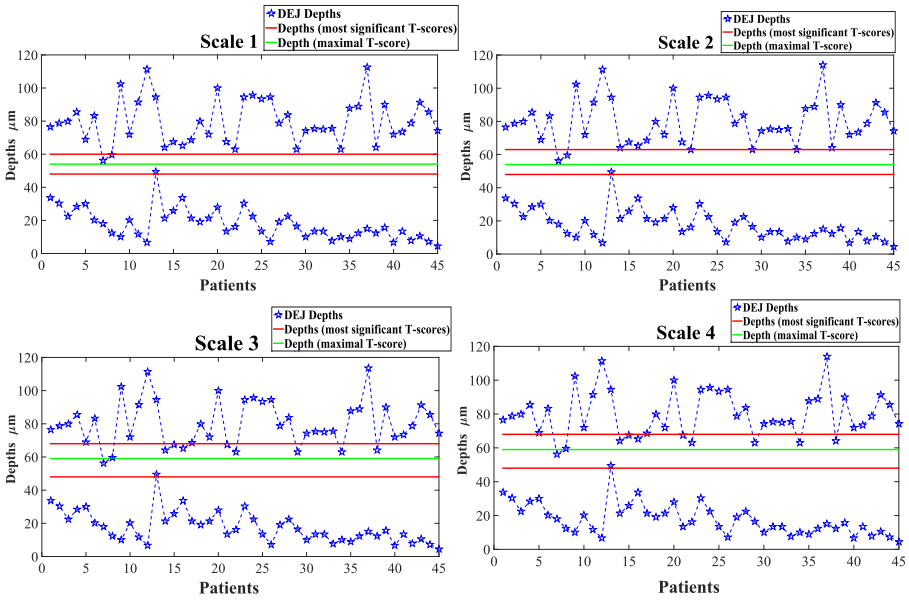

Fig. 10. Characteristics depths (found to be between 48um and 63um according to the T-test) and DEJ depths associated with the 45 patients.

Table 3. Confusion matrix of SVM classifiers based on Koller method using the same training and testing conditions as in Table.2. Features from images of the three characteristic depths have been concatenated in one feature vector. Slightly higher accuracy (77.7%) has been obtained when Koller's method is applied only to depth $54\mu m$.

	Koller method		
Confusion matrix	$\hat{L}$	$\hat{H}$	Sensitivity Specificity
Lentigo	21	6	77.7 %
Healthy	5	13	72.2 %
Precision	80.7 %	68.4 %	
Accuracy	75.5 %		

4. Discussion

The experimental results, with skin types 2 and 3, have shown that the lentigo phenomenon can be characterized using reflectance confocal microscopy images. Statistical tests demonstrated that it can be identified with sufficient statistical significance at depths between $50\mu m$ and $60\mu m$. Classification of patients with GGD statistical models associated with wavelet transforms of images acquired at these characteristic depths show that the parameter β can discriminate healthy and lentigo tissues. Other parameters including the scale parameter α, the variance and the entropy failed to achieve the same objective. Formally, $\beta \in \mathbb{R}^+$ is the shape parameter of the GGD distribution. When $\beta \to 0$, the GGD curve has a picked shape around the mode. However, when β is large, the curve has a rounded and more spread mode. Lower β indicate sparser wavelet coefficients. In our case, healthy images have values of β larger than those of lentigo (Fig.5). One possible interpretation of this result is related the visual structure of the dermal papillae in the lentigo images. In the case of lentigo, the papillae of the dermis grow higher and more dense towards the surface, with irregular shapes, making the DEJ junction larger. In our data, the DEJ starts at the mean depth of $13\mu m$ for lentigo and $25\mu m$ for healthy patients, with similar

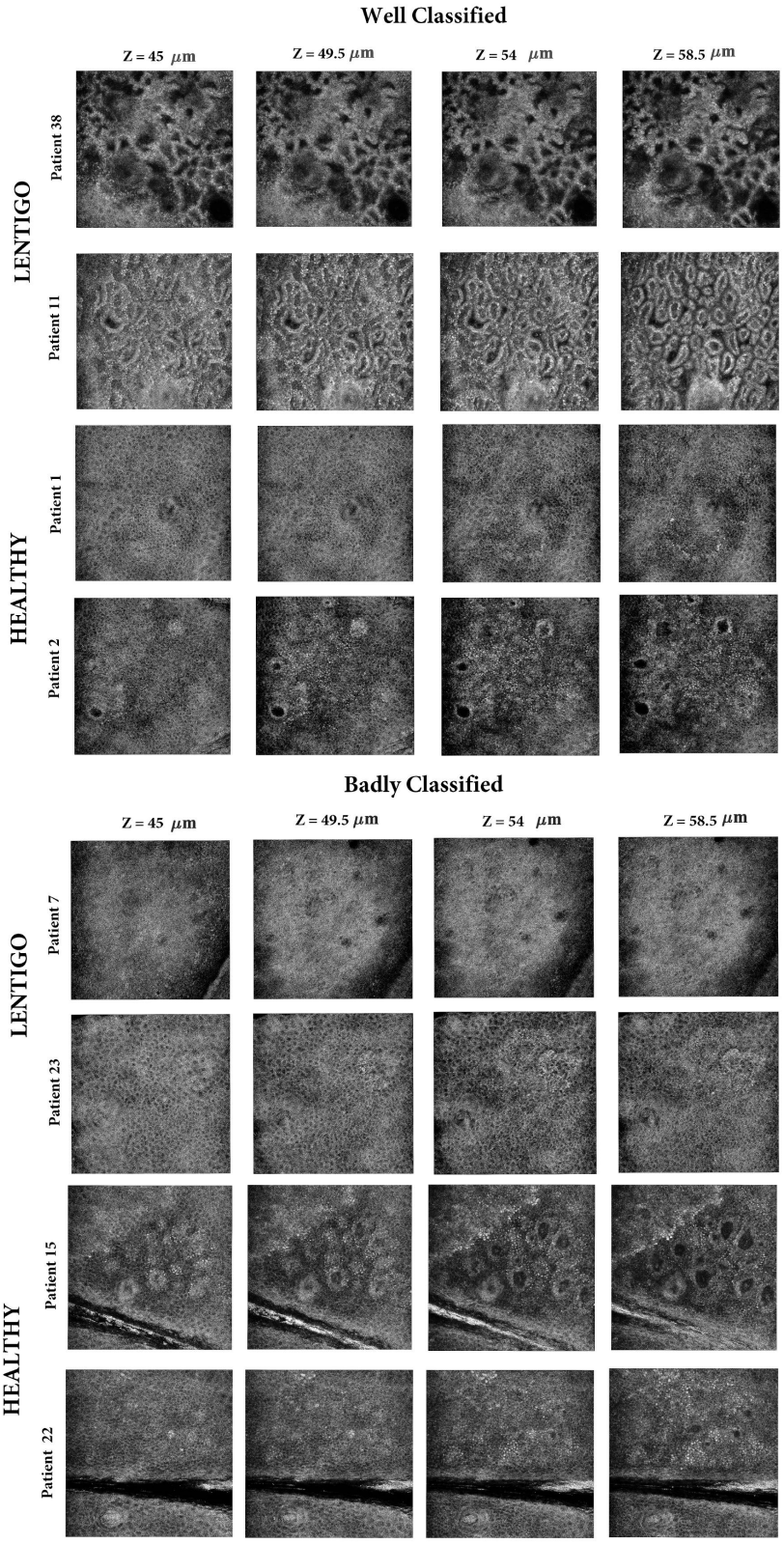

Fig. 11. Examples of RCM images of lentigo and healthy patients classified by the SVM classifier.

ending depths. As a consequence, images taken at the middle of the dermoepidermal junction contain more widespread and irregularly shaped papillae in lentigo than in healthy patients. The papillae appear as irregular shapes surrounded by bright borders that represent concentration of melanocytes (Fig. 2). These patterns of shapes are captured by the distribution of the wavelet coefficients as regular coarse textures with lower values of β. In comparison, the fewer and regular round shaped papillae that appear in the healthy images give slightly larger values of β. This numerical difference is statistically sufficient to distinguish healthy from lentigo patients. As shown in Fig.2, β decreases with the increase of the density and shape irregularity of the papillae in the image. Please notice that the value of β shown in this figure correspond to band H at scale #1. A similar behavior of this parameter is obtained for the other bands and scales. Finally, it is interesting to note that the proposed method would not discriminate lentigo from other skin lesions or types if their corresponding images have similar texture patterns, which is one of its limitations.

5. Conclusions

This paper investigated the potential of using the statistical properties of wavelet coefficients of RCM images to characterize the skin lentigo. The proposed method computed the scale and shape parameters of a generalized Gaussian distribution associated with each band and each scale of images acquired at different depths. These parameters were then used to distinguish between healthy and lentigo tissues. A parametric T-test was performed to assess the statistical significance of the observed differences between healthy and lentigo images. The proposed hypothesis test was run on a database of 2250 real images associated with 45 patients. SVM classifiers were created and tested for the different parameters. The results of t-test and SVM classification were found in perfect agreement. In conjunction, these results showed that the shape parameter β is well suited to discriminate healthy and lentigo tissues. They showed also that there is no preferred direction, and thus that it is better to use the joint vector $\boldsymbol{(H,V,D)}$ with the first scale to provide a good characterization of RCM images. Interestingly, for images from the back of the hand, this study found that the shape parameter β of the generalized Gaussian distribution characterizes lentigo at depths around $50\mu m$ corresponding to the location of the dermoepidermal junction. This result is in agreement with the clinical fact that lentigo significantly disorganizes this layer of the skin. Future work includes the consideration of different skin types and other pigmented skin lesions that lead to the destruction of the dermoepidermal junction such as the melanoma. It would be also interesting to analyze the distributions of RCM images directly in the image domain. The associated parameters could complement the statistical model defined by the scale and shape parameters of the generalized Gaussian distribution associated with the wavelet coefficients considered in this paper.

Funding

This work was funded by Pierre Fabre Dermo-cosmétique.

Disclosures

The authors declare that there are no conflicts of interest related to this article.