UNSUPERVISED BAYESIAN Model FOR THE CLASSIFICATION OF CUTANEOUS REFLECTANCE CONFOCAL MICROSCOPY IMAGES

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

[1]University of Toulouse, ENSEEIHT-IRIT, 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

ABSTRACT

This paper studies a new Bayesian algorithm for the classification of reflectance confocal microscopy (RCM) images of human skin. The objective of this algorithm is to identify the skin lentigo, a phenomenon that originates at the dermoepidermic junction (DEJ) from the healthy skin. The proposed Bayesian approach takes advantage of the distribution of the multiplicative speckle noise affecting these images and of appropriate priors for the unknown model parameters. A Metropolis-whitin-Gibbs sampler is then investigated to sample the posterior distribution of the Bayesian model associated with RCM images and to build estimators of its parameters, including labels indicating the class of each RCM image. The resulting algorithm is applied to synthetic data and to real images from a clinical study containing healthy and lentigo patients. The obtained classification performance is very encouraging.

1. INTRODUCTION

The lentigo is a hyperplasia that affects the skin. It comes from the proliferation of melanocyte cells at certain layers of the epidermis, mainly at the dermoepidermal junction, which leads to the disorganization of the regular cellular network[1]. Clinically, this disorder is assessed visually on the skin surface or through biopsy. Reflectance confocal microscopy (RCM) is a non-invasive imaging technique that is capable of capturing images of human skin from the epidermis to the papillary dermis in real time (down to the depth of 100 - 150 μm) [2]. This technique is also used to explore various skin lesions [3], including lentigo. For example, Fig. 1 shows examples of images from patients with and without lentigo (more images can be found in [4]). Various studies have attested of the usefulness of RCM for cancer and other tumor diagnosis [5]. Current practices to analyze these images are mainly based on visual inspection. In [1], the authors reported good correlation between RCM and histology in the case of melanoma. Studies of RCM has also proved valuable for treatment follow up [6], surveillance of lentigo malign treatment [7, 8], and guidance of cutaneous surgery [9].

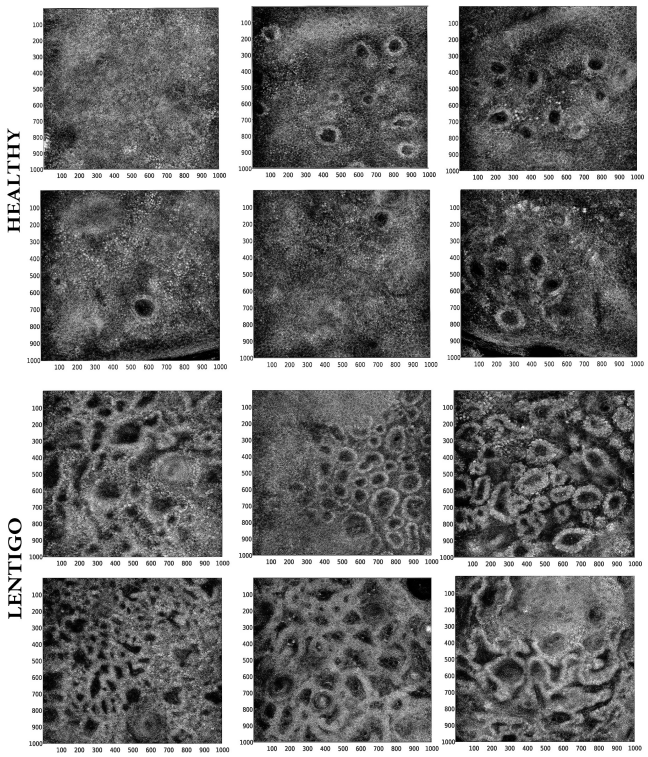


Fig. 1: Images (at the depth 54 μm) from healthy (patient #1, #2, #3, #4, #5, #6) and lentigo (patient #31, #33, #37, #38, #40, #44) patients at the DEJ depth. One can observe more textured images in the presence of lentigo.

However, RCM images are up to now mainly analyzed visually. Image processing methods could be helpful to exploit their potential and provide aid for medical decision making. Few of such methods were reported in the literature. In [10], Luck et al. developed a nuclei segmentation method based on a Gaussian model for the nuclei reflectivity and a truncated Gaussian distribution for the intensity of the cytoplasm fibers. Their Bayesian classification algorithm relies on a Gaussian

This work was funded by Pierre Fabre Dermo Cosmétique.

Markov random field exploiting spatial correlation between neighboring pixels of the analyzed images. Another application of RCM was developed and validated by Kurugol et al. to identify the dermoepidermal junction by classifying appropriate texture features [11, 12]. Hames et al. [13, 14] proposed a skin layer segmentation method for RCM images based on a logistic regression classifier. An SVM classification method was also developed in [15] to identify skin morphological patterns using RCM image texture features. Finally, a wavelet-based classification method was developed in [16] to distinguish benign and malignant melanocytic skin tumors. This method, which will be used as a benchmark in our study, is based on a decision tree classifier.

This paper studies a new Bayesian method for classifying RCM image pixels into two classes corresponding to healthy and lentigo tissues. Our first contribution is a hierarchical Bayesian model that allows a set of RCM images to be classified into healthy and lentigo classes. Each image is assumed to be corrupted by a multiplicative speckle noise with a gamma distribution. A truncated Gaussian distribution is then assigned to each image to classify, constraining these images to be positive. Prior distributions are finally assigned to the means and variances of these truncated Gaussian distributions, to the noise variances, and to the image labels. The joint posterior distribution of the proposed model is finally determined and will be used for image classification and parameter estimation. The second contribution of this paper is the derivation of an estimation algorithm associated with the proposed hierarchical Bayesian model. As the minimum mean square error (MMSE) and maximum a posteriori (MAP) estimators of the proposed model cannot be easily computed from its joint posterior, we investigate a hybrid Gibbs sampler allowing this posterior to be sampled (see [17, 18] for details). The proposed Bayesian model and estimation algorithm are validated using synthetic and real RCM images, resulting from a clinical study containing healthy and lentigo patients. The obtained results are very promising and show the potential of the proposed denoising and classification strategy.

The paper is structured as follows. The classification problem studied in this work is introduced in Section 2. The proposed hierarchical Bayesian model and its estimation algorithm are studied in Sections 3 and 4. Section 5.1 validates the proposed technique using simulated data with different noise levels. Section 5.2 shows results obtained using real data obtained from a clinical study. Conclusions and future work are finally reported in Section 6.

2. PROBLEM FORMULATION

2.1. Observation model

Consider L noise free images, containing N pixels, gathered in the matrix $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_l] \in \mathbb{R}^{N\times L}$, where $\boldsymbol{s}_l$, $l \in \{1, \cdots, L\}$ denotes the image associated with the lth patient. Denote by $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l] \in \mathbb{R}^{N\times L}$ the corresponding noisy images. Using these notations, the observation model is given by

$$\boldsymbol{y}_l = \boldsymbol{s}_l \odot \boldsymbol{b}_l, \text{ with } \boldsymbol{b}_l \sim \mathcal{G}(\rho_l, \theta_l) \tag{1}$$

where $\boldsymbol{y}_l$ and $\boldsymbol{s}_l$ are $(N \times 1)$ vectors representing the lth observed and noiseless images, $\boldsymbol{b}_l$ is a gamma noise $(N \times 1)$ vector with a shape parameter ρ_l and a scale parameter θ_l and $\odot$ denotes the termwise product. In order to ensure that the proposed model (1) is identifiable, the mean of the gamma noise is supposed to equal 1, leading to

$$\mathbb{E}(\boldsymbol{b}_l) = 1 \quad \Rightarrow \quad \rho_l = \frac{1}{\theta_l}. \tag{2}$$

The problem addressed in this paper is to classify these images $\boldsymbol{y}_l, l \in \{1, \cdots, L\}$ into two classes representing healthy and lentigo patients. The next section introduces a hierarchical Bayesian model that is used for this classification.

3. HIERARCHICAL BAYESIAN MODEL

This section introduces a hierarchical Bayesian model that can be used to estimate the unknown $N \times L$ matrix of noiseless images $\boldsymbol{S}$, the $L \times 1$ vectors $(\boldsymbol{z}, \boldsymbol{\theta})$ containing the class labels and the noise variances associated with the L observed images from the matrix $\boldsymbol{Y}$. This model is defined by a likelihood, and by parameter and hyperparameter priors defined below.

3.1. Likelihood

The multiplicative speckle noise $\boldsymbol{b}_l$ is known to have a gamma distribution. Thus, the observation model (1) leads to

$$y_{nl}|s_{nl}, \theta_l \sim \mathcal{G}\left(\frac{1}{\theta_l}, s_{nl}\,\theta_l\right) \tag{3}$$

where $\sim$ means "is distributed according to", $\mathcal{G}$ is the gamma distribution whose probability density function (pdf) is

$$f(y_{nl} \mid s_{nl}, \theta_l) \propto \frac{(y_{nl})^{\frac{1}{\theta_l}-1} \exp\left(-\frac{y_{nl}}{s_{nl}\,\theta_l}\right)}{\Gamma\left(\frac{1}{\theta_l}\right)(s_{nl}\,\theta_l)^{\frac{1}{\theta_l}}} I_{\mathbb{R}^+}(y_{nl}) \tag{4}$$

with $I_{\mathbb{R}^+}(y_{nl})$ the indicator function on $\mathbb{R}^+$, $\propto$ means "proportional to" and Γ denotes the gamma function. Assuming independence between the observed signals, the likelihood of the L observed images can be written

$$f(\boldsymbol{Y}|\boldsymbol{S}, \boldsymbol{\theta}) \propto \prod_{n=1}^{N}\prod_{l=1}^{L} f(y_{nl}|s_{nl}, \theta_l).$$

3.2. Priors for the signal of interest

To ensure the positivity of the noiseless images, a truncated Gaussian distribution is assigned to $\boldsymbol{s}_l$ for $l \in \{1, \cdots, L\}$

$$\boldsymbol{s}_l \mid z_l = k, \mu_k, \sigma_k^2 \sim \mathcal{N}_{\mathbb{R}^+}(\mu_k, \sigma_k^2) \tag{5}$$

where $\mathcal{N}_{\mathcal{S}}$ denotes the truncated normal distribution on $\mathcal{S}$, k takes the two values 1 and 2 depending on the patient class, and (μ_k, σ_k^2) are the means and variance of the two truncated Gaussian distributions.

3.3. Prior for the noise variances

A non-informative conjugate inverse gamma prior (denoted as $\mathcal{IG}$) is classically selected for the scale parameter θ_j [19]

$$\theta_l \mid a, b \sim \mathcal{IG}(a, b) \tag{6}$$

where a and b are fixed hyperparameters, that are adjusted to reflect the absence of prior knowledge on θ_l, i.e., the mean and variance of θ_l were fixed to 1 and 100 in order to obtain a flat prior. The joint prior for the vector of noise variances denoted as $f(\boldsymbol{\theta} \mid a, b)$ is finally obtained as the product of the marginal densities $f(\theta_i \mid a, b)$.

3.4. Prior for the label vector $\boldsymbol{z}$

The parameter vector $\boldsymbol{z} = (z_1, ..., z_L)$ is a label vector that associates each image to a healthy or lentigo skin. Because of the absence of prior knowledge about this parameter, it is assigned a uniform prior defined as

$$P(z_l = k) = \frac{1}{2}, \forall l = 1, ..., L. \tag{7}$$

The labels associated with the different patients are supposed to be a priori independent, i.e., the joint prior of $\boldsymbol{z}$ denoted as f($\boldsymbol{z}$) is the product of the probabilities defined in (7).

3.5. Hyperparameter priors

In order to complete the description of the proposed hierarchical Bayesian model and to allow hyperparameters to be estimated directly from the data, we propose to assign priors for the different hyperparameters. A Gaussian prior has been selected for the mean μ_k and a non-informative inverse gamma prior for the variance σ_k^2 (see [19,20] for motivations)

$$\mu_k \mid \mu_0, \sigma_0 \sim \mathcal{N}(\mu_0, \sigma_0^2) \tag{8}$$

$$\sigma_k^2 \mid \alpha_0, \beta_0 \sim \mathcal{IG}(\alpha_0, \beta_0) \tag{9}$$

where μ_0, σ_0^2, α_0, β_0 are fixed in order to obtain flat priors, i.e., $\mu_0 = 100$, $\sigma_0^2 = 10^5$ whereas the mean and variance of σ_k^2 were fixed to 1 and 1000. The joint pdfs $f(\boldsymbol{\mu} \mid \mu_0, \sigma_0)$ and $f(\boldsymbol{\sigma}^2 \mid \alpha_0, \beta_0)$ are finally obtained as the product of their marginal densities assuming prior independency between the components of these two vectors.

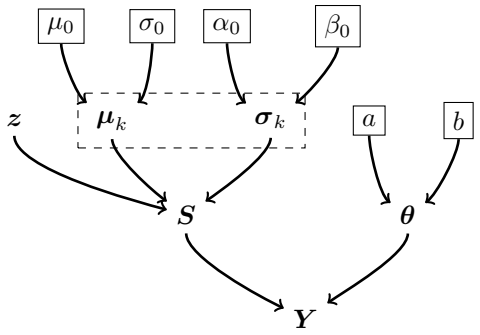


Fig. 2: DAG for the parameter and hyperparameter priors. The user fixed hyperparameters appear in boxes (continuous line).

3.6. Joint posterior distribution

The proposed Bayesian model is illustrated by the directed acyclic graph (DAG) displayed in Fig. 2, which highlights the relationships between the observations $\boldsymbol{Y}$, the parameters $\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{z}$ and the hyperparameters μ_k, σ_k^2. Assuming prior independence between the different components of the parameter vector $\boldsymbol{X} = \left(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{z}, \mu_k, \sigma_k^2\right)$, the joint posterior distribution of this Bayesian model can be computed using the following hierarchical structure

$$f(\boldsymbol{X} \mid \boldsymbol{Y}) \propto f(\boldsymbol{Y} \mid \boldsymbol{S}, \boldsymbol{\theta}) f(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) \tag{10}$$

$$\begin{aligned} \text{with } f(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) &= f(\boldsymbol{S} \mid \boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) f(\boldsymbol{\theta} \mid a, b) \\ &\times f(\boldsymbol{\mu} \mid \mu_0, \sigma_0) f(\boldsymbol{\sigma}^2 \mid \alpha_0, \beta_0) f(\boldsymbol{z}). \end{aligned} \tag{11}$$

The complexity of the proposed Bayesian model summarized in the DAG of Fig. 2 and its resulting posterior (10) prevent a simple computation of the maximum a-posteriori (MAP) or minimum mean square (MSE) estimators of the unknown model parameters. The next section studies a Markov chain Monte Carlo (MCMC) method that is used to sample the posterior (10) and to build estimators of the parameters involved in the proposed Bayesian model using the generated samples.

4. METROPOLIS-WITHIN-GIBBS ALGORITHM

This section studies a hybrid-Gibbs-sampler, which is guaranteed to generate samples asymptotically distributed according to the target distribution (10). The Gibbs sampler described in Algo. 1, iteratively generates samples distributed according to the conditional distributions of (10) that are not described here for brevity (see [4, Appendix A] for more details regarding these distributions). Because of the complexity of the conditional distributions, we consider random-walk Metropolis-Hastings (RWMH) [17, 18] moves within the Gibbs sampler, which requires the definition of proposal distributions for each conditional distribution that is

not easy to sample. In our case, we use a truncated Gaussian as a proposal distribution for the parameters $\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2$ and a Gaussian distribution for $\boldsymbol{\mu}$. The main steps of the proposed

Algorithm 1 Metropolis-within-Gibbs algorithm

Input: $\mathrm{N_{bi}}, \mathrm{N_{MC}}, \boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2$
Initialization
Initialize $\boldsymbol{S}^{(0)}, \boldsymbol{\theta}^{(0)}, \boldsymbol{z}^{(0)}, \boldsymbol{\mu}^{(0)}, \boldsymbol{\sigma}^{2(0)}$
for i=1 to N_{MC} **do**
  Parameter update
  Sample $\boldsymbol{S}^{(i)} \mid \boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2$ according to (20) in 8 using an RWMH with a truncated Gaussian proposal
  Sample $\boldsymbol{\theta}^{(i)} \mid \boldsymbol{Y}, \boldsymbol{S}, a, b$ according to (21) in 8 using an RWMH with a truncated Gaussian proposal
  Sample $\boldsymbol{\mu}^{(i)} \mid \boldsymbol{S}, \boldsymbol{\sigma}^2, \mu_0, \sigma_0^2$ according to (22) in 8 using an RWMH with a Gaussian proposal
  Sample $\boldsymbol{\sigma}^{2(i)} \mid \boldsymbol{S}, \boldsymbol{\mu}, \alpha_0, \beta_0$ according to (23) in 8 using an RWMH with a truncated Gaussian proposal
  Sample $\boldsymbol{z}^{(i)} \mid \boldsymbol{S}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2$ from the pdf (24) in 8
end for
Result: $\boldsymbol{S}^{(i)}, \boldsymbol{\theta}^{(i)}, \boldsymbol{z}^{(i)}, \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{2(i)}$ for $i = 1, ..., \mathrm{N_{MC}}$.

Metropolis-within-Gibbs sampler are summarized in Algo. 1. This algorithm provides a sequence of samples of the vector $\boldsymbol{X} = \left(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{z}, \mu_k, \sigma_k^2\right)$ denoted as $\boldsymbol{X}_j^{(i)}$ that are used to approximate the MMSE estimators by using Monte-Carlo integration [21] as

$$\boldsymbol{X}^{\mathrm{MMSE}} \simeq \frac{1}{\mathrm{N_{MC}} - \mathrm{N_{bi}}} \sum_{i=\mathrm{N_{bi}}+1}^{\mathrm{N_{MC}}} \boldsymbol{X}^{(i)} \tag{12}$$

where $\mathrm{N_{bi}}$ is the number of burn-in iterations and $\mathrm{N_{MC}}$ is the total number of Monte Calo iterations. Finally, the following maximum a-posteriori (MAP) estimator is considered for the label $\boldsymbol{z}$

$$\boldsymbol{z}_l^{\mathrm{MAP}} \simeq \begin{cases} 1 & \text{if } \left[\boldsymbol{z}_l^{(i)} = 1\right]_{i=\mathrm{N_{bi}}+1}^{\mathrm{N_{MC}}} \geq \left[\boldsymbol{z}_l^{(i)} = 2\right]_{i=\mathrm{N_{bi}}+1}^{\mathrm{N_{MC}}} \\ 2 & \text{otherwise} \end{cases} \tag{13}$$

where $[x = 1]_i^j$ and $[x = 2]_i^j$ denote the numbers of samples satisfying the conditions $x = 1$ and $x = 2$ in the interval $[i, j]$.

4.1. Convergence:

Running multiple chains with different initializations allows to define various convergence measures for MCMC methods [22]. The popular between-within variance criterion has shown interesting properties for diagnosing convergence of MCMC methods. This criterion was initially studied by Gelman and Rubin in [23] and has been used in many studies including [22, p. 33], [24] and [25]. The main idea is to run M parallel chains of length $\mathrm{N_r}$ + $\mathrm{N_{bi}}$ for each data set with different starting values and to evaluate the dispersion of the estimates obtained from the different chains. The between-sequence variance B and within-sequence variance W for the M Markov chains are defined by

$$B = \frac{\mathrm{N_r}}{M-1} \sum_{m=1}^{M} \left(\bar{k}_m - \bar{k}\right)^2 \tag{14}$$

$$W = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{\mathrm{N_r}} \sum_{t=1}^{N_r} \left(k_m^{(t)} - \bar{k}_m\right)^2 \tag{15}$$

with

$$\bar{k}_m = \frac{1}{\mathrm{N_r}} \sum_{t=1}^{\mathrm{N_r}} k_m^{(t)}, \;\; \bar{k} = \frac{1}{M} \sum_{m=1}^{M} \bar{k}_m, \;\; \mathrm{N_r} = \mathrm{N_{MC}} - \mathrm{N_{bi}}. \tag{16}$$

where k is the parameter of interest and $k_m^{(t)}$ is its estimate at the tth run of mth chain. The convergence of the chain can then be monitored by the so-called *potential scale reduction factor* $\hat{\rho}$ defined as [26, p. 332]

$$\sqrt{\hat{\rho}} = \sqrt{\frac{1}{W}\left(\frac{\mathrm{N_r} - 1}{\mathrm{N_r}} W + \frac{1}{\mathrm{N_r}} B\right)}. \tag{17}$$

5. SIMULATION RESULTS

5.1. Synthetic data

This section evaluates the performance of the proposed algorithm on synthetic data. Different experiments were conducted using three values of the signal to noise ratio SNR $\in \{0\text{ dB}, 10\text{ dB}, 20\text{ dB}\}$, allowing the algorithm performance to be appreciated for different noise levels. This section considers $L = 100$ synthetic images. Each image contains $N = 2000$ pixels and was generated according to (3). These images were separated into healthy and lentigo classes containing 50 images. The noiseless images of the two classes were respectively generated according to the truncated Gaussian distributions $\mathcal{N}_{\mathbb{R}^+}(\mu_1, \sigma_1^2)$ and $\mathcal{N}_{\mathbb{R}^+}(\mu_2, \sigma_2^2)$, with $\mu_1 = 17, \mu_2 = 20, \sigma_1^2 = 2, \sigma_2^2 = 4$. The sampler convergence of the algorithm is monitored by computing the potential scale reduction factor introduced in (4.1) for an appropriate parameter of interest. Different choices for the parameter k could be considered for the proposed method. This paper proposes to monitor the convergence of the Metropolis-within-Gibbs sampler by checking the noise variance θ (see [20, 24] for a similar choice). The potential scale reduction factor for parameter θ computed for $M = 10$ Markov chains is equal to 1.01. This value of $\sqrt{\hat{\rho}}$ confirms the good convergence of the sampler (a recommendation for convergence assessment is a value of $\sqrt{\hat{\rho}} \leq 1.2$ [26, p. 332]). Figs. 3, 4 and 5 show the evolution of the Markov chains for the different

parameters $\hat{\mu_1}, \hat{\mu_2}, \hat{\sigma_1}, \hat{\sigma_2}, \boldsymbol{\theta}$ estimated for synthetic data with $\text{SNR}_Y = [0 \text{ dB}, 10 \text{ dB}, 20 \text{ dB}]$, respectively. Algo. 1 was run for $\text{N}_{\text{MC}} = 100000$ iterations and the different model parameters were estimated using (12) and (13) using a burn-in period of length $\text{N}_{\text{bi}} = 99900$. The performance of the algorithm was evaluated by computing the mean square errors (MSEs) of the different parameters and the signal to noise ratios (SNRs) defined as

$$\text{MSE}_j = \| \hat{\boldsymbol{X}}_j - \boldsymbol{X}_j \|^2 \tag{18}$$

$$\text{SNR}_j = 20 \log_{10} \left(\frac{||\boldsymbol{X}_j||}{||\boldsymbol{X}_j - \widehat{\boldsymbol{X}}_j||} \right). \tag{19}$$

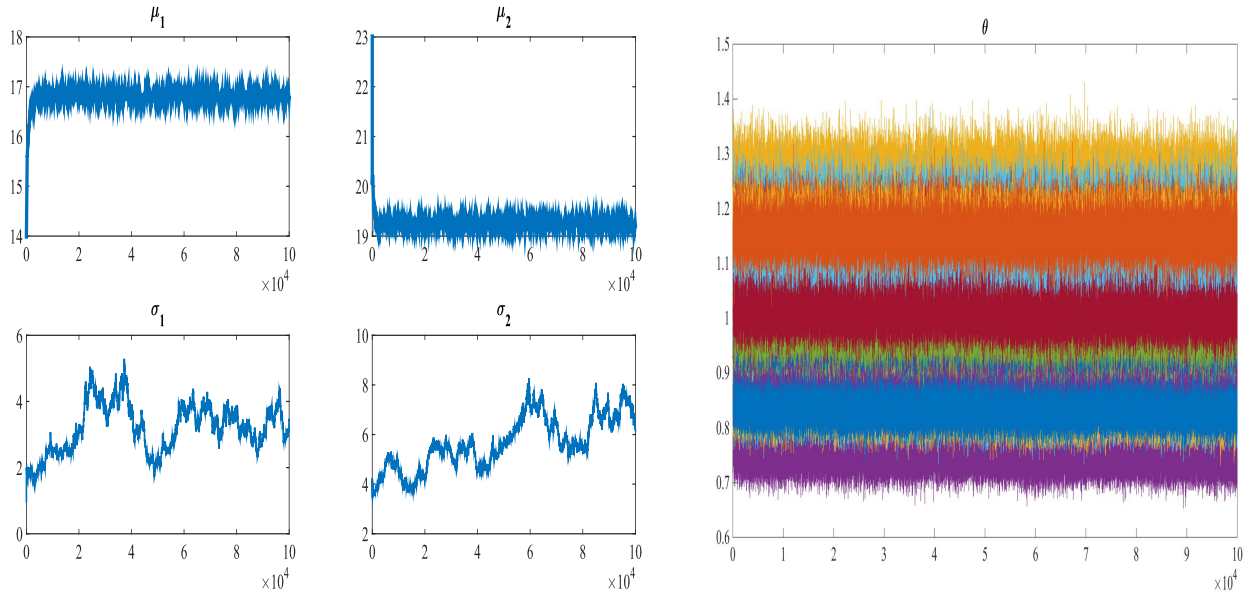
Quantitative results are presented in Table 1 for the three experiments. This table shows good estimation results of the parameters when considering different noise levels. The Table also shows excellent classification results for $\text{SNR}_Y \geq 10$ dB, and 91% when considering the challenging case $\text{SNR}_Y = 0$ dB. These results highlight the potential of the proposed strategy in denoising and classifying the images obtained from model (3) and improving the estimation of the different parameters of this model.

5.2. Real data

This section is devoted to the validation of the proposed denoising and classification algorithm when applied to real RCM images. These RCM images were acquired with apparatus Vivascope 1500 and correspond to 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 with 1000×1000 pixels. A set of $L = 45$ women aged 60 years and over were recruited. 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. The first group was formed by 27 women with at least 3 lentigines on the back of the hand whereas 18 women without lentigo constituted the control group. 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 in each image. Given the large size of the images, we preferred to select and apply our algorithm to patches of 250×250 pixels for each image to reduce the computational cost. Fig. 6 show the evolution of the Markov chains convergence of the different estimated parameters $(\hat{\mu_1}, \hat{\mu_2}, \hat{\sigma_1}, \hat{\sigma_2}, \boldsymbol{\theta})$ for these RCM images. The obtained results were then used to calculate the confusion matrix and four indicators (sensitivity, specificity, precision, accuracy) shown in Tables (2) and (3). These indicators are defined as Sensitivity = TP/(TP+FN), Specificity = TN/(FP+TN), Precision = TP/(TP+FP), Accuracy = (TP+TN)/(TP+FN+FP+TN), where TP, TN, FP and FN are the numbers of true positives, true negatives, false positives and false negatives. The Table (2) allows us to evaluate the classification performance of the proposed strategy. The accuracy of the proposed method equals 97.7%, which corresponds to a single mistake for the lentigo patient $\#8$. Fig. 7 shows that the texture of this misclassified image is not very destructed as for other lentigo patients, and is visually similar to the texture of healthy patients. Fig. 8 shows examples of noisy RCM images and their estimated true reflectivity. We can observe that the estimated images have low intensity compared to the noisy images which is due to the fact that the noise is multiplicative. To assess the significance of our results, our algorithm was then compared to the method presented in [16]. This method consists in extracting from each RCM image a set of 39 analysis parameters (further technical details are available in [27]) and to apply to these features a classification procedure performed by the CART (Classification and Regression Trees). The latter tested on the considered real RCM images after training using a leave one out procedure. As shown in Table (3), the obtained accuracy was 82.2% confirming the good performance of the proposed classifier.

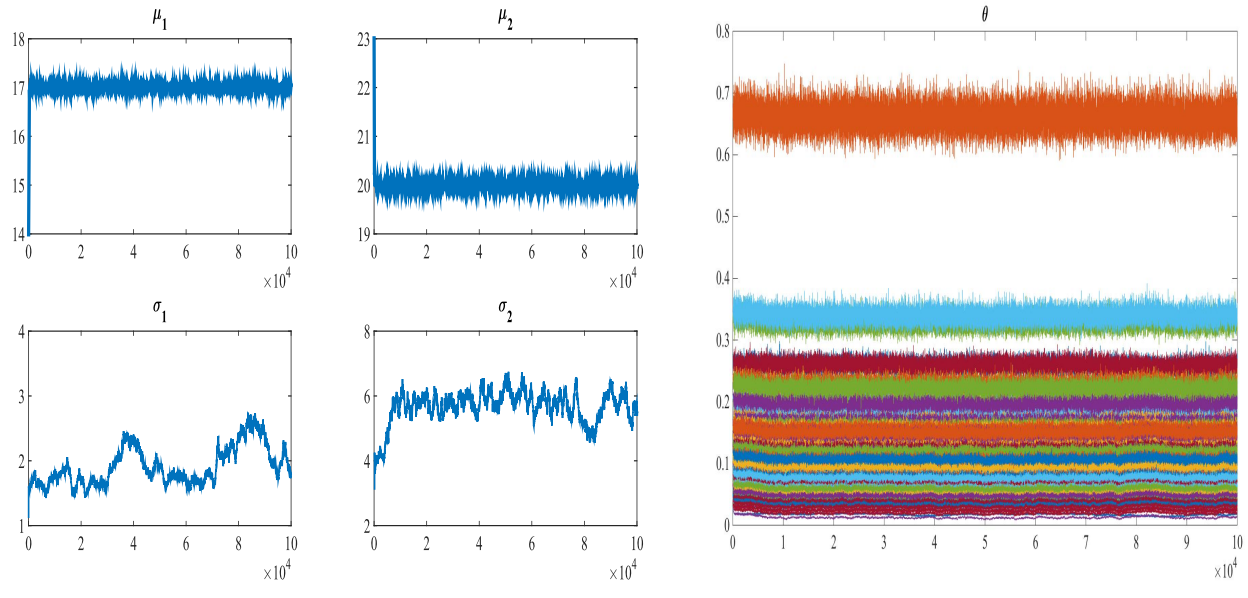
6. CONCLUSIONS

This paper presented a new unsupervised hierarchical Bayesian strategy for reconstruction and classification of RCM images as healthy or lentigo patients. The proposed Bayesian model was introduced based on a gamma distribution for the multiplicative speckle noise and on various priors assigned to the unknown model parameters in order to regularize the estimation problem. An MCMC algorithm was then proposed to jointly estimate the different parameters of the model, including the true reflectivity of the images, the label vector allowing the classification of the images in the classes "healthy" or "lentigo" and finally the parameters of the speckle noise. More precisely, a Metropolis-within-Gibbs sampler is used to generate samples distributed asymptotically according to the conditional posterior of this Bayesian model. These samples are then used to construct Bayesian estimators, such as the maximum a posteriori estimator (MAP) or the minimum mean square error (MMSE) estimator for the unknown parameters of the model. The resulting algorithm was then compared to the CART method. Simulation results conducted on synthetic and real data allowed the good performance of the proposed classifier to be appreciated. Future work includes the introduction of spatial correlation on the estimated noiseless images to improve their quality.



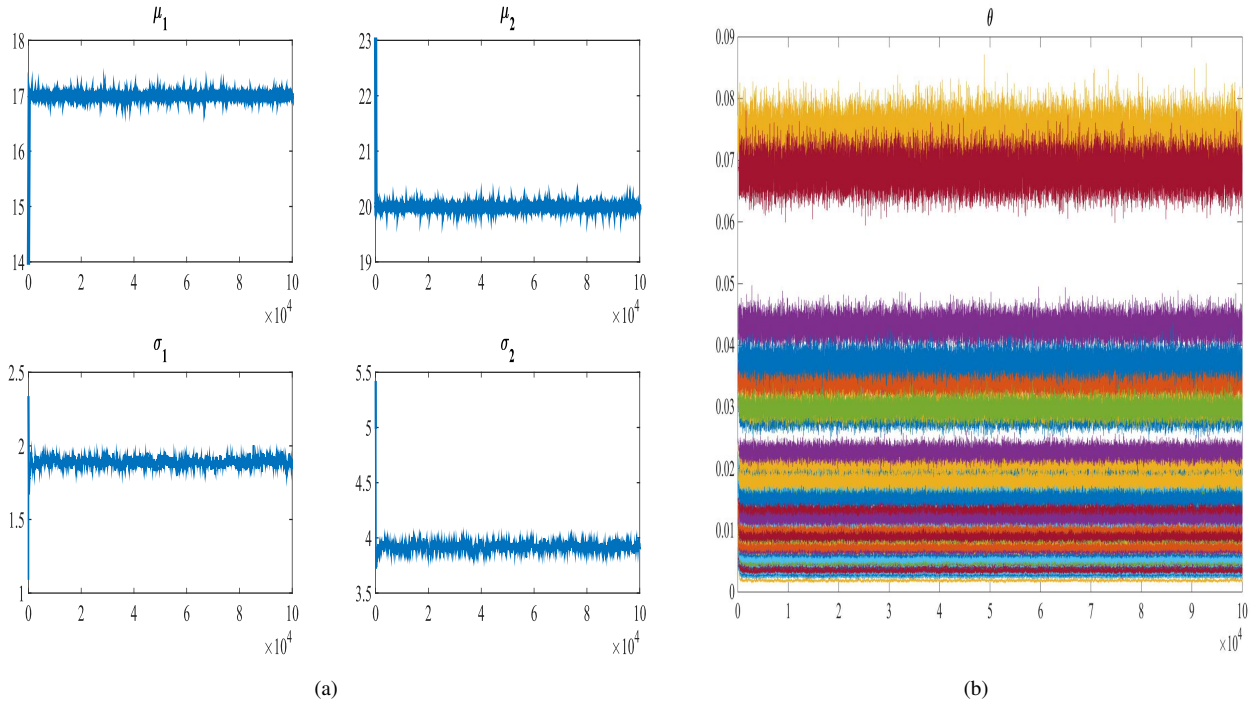
(a) (b)

Fig. 3: Evolution of the convergence of the Markov chains for the different parameters $\hat{\mu_1}, \hat{\mu_2}, \hat{\sigma_1}, \hat{\sigma_2}, \boldsymbol{\theta}$ estimated for the synthetic data with $\text{SNR}_Y = 0$ dB.



(a) (b)

Fig. 4: Evolution of the convergence of the Markov chains for the different parameters $\hat{\mu_1}, \hat{\mu_2}, \hat{\sigma_1}, \hat{\sigma_2}, \boldsymbol{\theta}$ estimated for the synthetic data with $\text{SNR}_Y = 10$ dB.

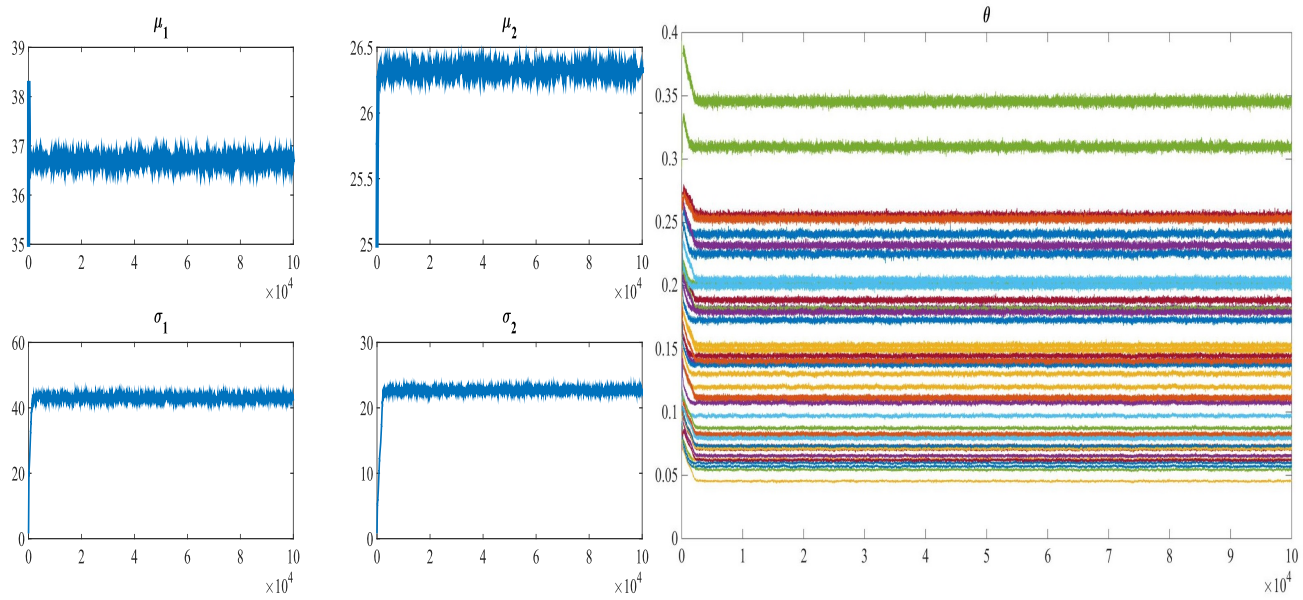


(a) (b)

Fig. 5: Evolution of the convergence of the Markov chains for the different parameters $\hat{\mu_1}, \hat{\mu_2}, \hat{\sigma_1}, \hat{\sigma_2}, \boldsymbol{\theta}$ estimated for the synthetic data with $\text{SNR}_Y = 20$ dB.

Table 1: Performance of the proposed algorithm for denoising and classification of synthetic data for three corrupted data $\text{SNR}_Y = [0 \text{ dB}, 10 \text{ dB}, 20 \text{ dB}]$.

	$\text{SNR}_Y = 0$ dB		$\text{SNR}_Y = 10$ dB		$\text{SNR}_Y = 20$ dB	
	MSE^2	SNR (dB)	MSE^2	SNR (dB)	MSE^2	SNR (dB)
μ_1	0.56	30.12	$1.54.10^{-4}$	62.72	$2.63.10^{-5}$	70.4
μ_2	0.95	21.42	$1.89.10^{-5}$	73.24	$6.64.10^{-5}$	67.79
σ_1^2	2.91	1.01	0.015	18.07	0.011	25.7
σ_2^2	7.14	2.57	4.58	5.42	0.006	22.07
$\boldsymbol{\theta}$	$1.14.10^{-3}$	20.44	$4.74.10^{-5}$	26.56	$5.68.10^{-7}$	30.44
$\boldsymbol{S}$	5.48	16.53	2.88	20.81	0.7093	26.87
Accuracy	91%		100%		100%	
Accuracy (CART)	83%		100%		100%	



(a) (b)

Fig. 6: Evolution of the convergence of the Markov chaines for the different parameters $\hat{\mu_1}, \hat{\mu_2}, \hat{\sigma_1}, \hat{\sigma_2}, \boldsymbol{\theta}$ estimated for the real RCM images.

Table 2: Classification performance on real data (45 patients) using the proposed method.

Confusion matrix	$\widehat{\mathbf{L}}$	$\widehat{\mathbf{H}}$	**Sensitivity** **Specificity**
Lentigo	**26**	**1**	**96.2 %**
Healthy	**0**	**18**	**100 %**
Precision	**100 %**	**94.7 %**	
Accuracy	**97.7 %**		

Table 3: Classification performance on real data (45 patients) using the CART method.

Confusion matrix	$\widehat{\mathbf{L}}$	$\widehat{\mathbf{H}}$	**Sensitivity** **Specificity**
Lentigo	**24**	**3**	**88.8 %**
Healthy	**5**	**13**	**72.2 %**
Precision	**82.7 %**	**81.2 %**	
Accuracy	**82.2 %**		

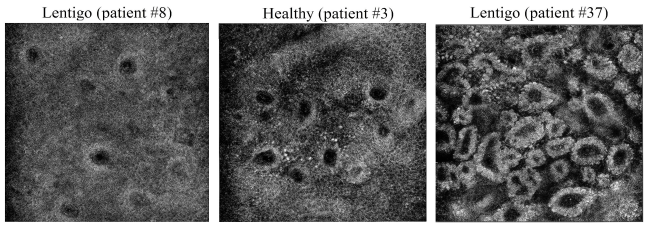


Fig. 7: Images from the patient #8 who is badly classified compared to a healthy and lentigo patient (well classified). One can observe more similarity between this patient and the healthy one then with the lentigo.

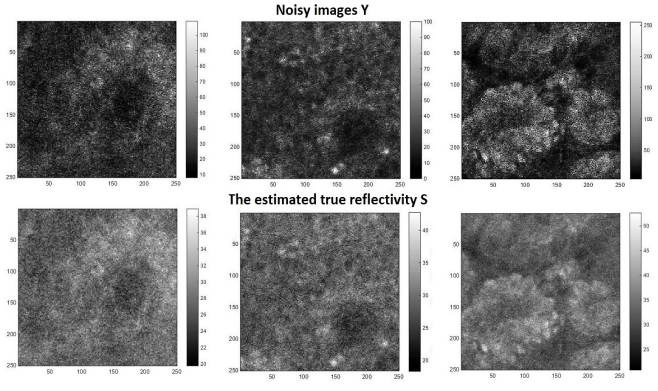


Fig. 8: Examples of noisy images (at the depth 54 μm) and their estimated true reflectivities.

7. APPENDIX: THE RANDOM-WALK METROPOLIS-HASTINGS

The RWMH used in Algo. 1, consists in generating samples distributed according to the complex conditional distribution of each parameter of interest. This is achieved using the "J" conditional distributions $f_j(.)$, for $j \in 1, ..., J$, and their associated proposal distributions $g_j(.), \forall j$. The first step is to initialize the sample value for each parameter $\boldsymbol{X}_j^{(0)}$, for $j \in 1, ..., J$. The main loop of the RWMH algorithm consists of three components:

1. Generate a proposal (or a candidate) sample $\boldsymbol{X}_j^{\text{cand}}$ from the proposal distribution $g_j\left(\boldsymbol{X}_j^{(\text{cand})} \mid \boldsymbol{X}_j^{(i-1)}\right)$ which is the truncated Gaussian distribution $\mathcal{N}_{\mathbb{R}^+}\left(\boldsymbol{X}_j^{(i-1)}, \boldsymbol{\epsilon_j^2}\right)$ (generated using [28]) for the parameters $\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2$ and a Gaussian distribution $\mathcal{N}\left(\boldsymbol{X}_{\boldsymbol{\mu}}^{(i-1)}, \boldsymbol{\epsilon_\mu^2}\right)$ for $\boldsymbol{\mu}$.
2. Compute the acceptance probability via the acceptance function $\alpha\left(\boldsymbol{X}_j^{(\text{cand})} \mid \boldsymbol{X}_j^{(i-1)}\right)$ based upon the proposal distribution and the conditional density for each parameter
   where $\alpha = \min\left\{\frac{f_j\left(\boldsymbol{X}_j^{(\text{cand})}\right)}{f_j\left(\boldsymbol{X}_j^{(i-1)}\right)} \; \frac{g_j\left(\boldsymbol{X}_j^{(i-1)} | \boldsymbol{X}_j^{(\text{cand})}\right)}{g_j\left(\boldsymbol{X}_j^{(\text{cand})} | \boldsymbol{X}_j^{(i-1)}\right)}, 1\right\}$
3. Accept the candidate sample with probability α.

In order to maximise the efficiency of the algorithm, the variances of the proposal distributions $\boldsymbol{\epsilon_j^2}$, have been adjusted such that the acceptance rate is between 0.3 and 0.6 as suggested in [17, 29].

8. APPENDIX: SAMPLING THE CONDITIONAL DISTRIBUTIONS

The conditional distributions are obtained by multiplying the likelihood with the different priors and by removing the multiplicative terms that do not depend on the variable of interest. The algorithm iteratively updates each parameter by using its conditional distribution. The results are detailed in the following paragraphs:

8.1. Sampling the parameter s_{nl}

We obtain the following conditional law

$$f\left(s_{nl} \mid z_l = k, y_{nl}, \theta_l, \sigma_k^2, \mu_k\right) \propto \frac{1}{\left(s_{nl}\right)^{1/\theta}} \; \exp\left(-\frac{y_{nl}}{s_{nl}\,\theta_l}\right)$$

$$\times \; \exp\left[-\frac{1}{2\,\sigma_k^2}\left(s_{nl} - \mu_k\right)^2\right] I_{\mathbb{R}^+}(s_{nl}) \qquad (20)$$

Using the following proposal law (positive truncated Gaussian [30])

$$g\left(x \mid s_{nl}^t\right) \propto \frac{\frac{1}{\sqrt{2\,\pi}\,\epsilon_2} \; \exp\left[-\frac{(x - s_{nl}^t)^2}{2\,\epsilon_2^2}\right]}{1 - \Phi\left(\frac{-s_{nl}^t}{\epsilon_2}\right)} \quad ,$$

where Φ is the cumulative normal distribution function
We obtain the following acceptance-rejection rule

$$s_{nl}^{t+1} = \begin{cases} s_{nl}^* & \text{with prob } \min\left\{\frac{f(s_{nl}^*)}{f(s_{nl}^t)} \; \frac{g(s_{nl}^t | s_{nl}^*)}{g(s_{nl}^* | s_{nl}^t)}, 1\right\} \\ s_{nl}^t & else \end{cases}$$

with

$$\frac{f(s_{nl}^*)}{f(s_{nl}^t)} \; \frac{g(s_{nl}^t \mid s_{nl}^*)}{g(s_{nl}^* \mid s_{nl}^t)} = \left(\frac{s_{nl}^*}{s_{nl}^t}\right)^{-\rho_l} \; \left[\frac{1 - \Phi\left(\frac{-s_{nl}^t}{\epsilon_2}\right)}{1 - \Phi\left(\frac{-s_{nl}^*}{\epsilon_2}\right)}\right] \times$$

$$\exp\left[\frac{-2\,y_{nl}\,\sigma_k^2 - s_{nl}^*\,\theta_l\,(s_{nl}^* - \mu_k)^2}{2\,\sigma_k^2\,s_{nl}^*\,\theta_l} + \frac{2\,y_{nl}\,\sigma_k^2 + s_{nl}^t\,\theta_l\,(s_{nl}^t - \mu_k)^2}{2\,\sigma_k^2\,s_{nl}^t\,\theta_l}\right].$$

8.2. Sampling the parameter θ_l

We obtain the following conditional law

$$f\left(\theta_l \mid y_{nl}, s_{nl}, a, b\right) \propto \frac{1}{{\theta_l}^{N/\theta_l + a + 1}} \; \prod_{n=1}^{N} \left(\frac{y_{nl}}{s_{nl}}\right)^{\frac{1}{\theta_l}}$$

$$\times \; \exp\left[-\frac{1}{\theta_l}\left(\sum_{n=1}^{N} \frac{y_{nl}}{s_{nl}} - b\right)\right] \; \left[\Gamma\left(1/\theta_l\right)\right]^{-N} \qquad (21)$$

Using the following proposal law (positive truncated Gaussian [30])

$$g\left(x \mid \theta_l^t\right) \propto \frac{\frac{1}{\sqrt{2\,\pi}\,\epsilon_1} \; \exp\left[-\frac{(x - \theta_l^t)^2}{2\,\epsilon_1^2}\right]}{1 - \Phi\left(\frac{-\theta_l^t}{\epsilon_1}\right)}.$$

where Φ is the cumulative normal distribution function
We obtain the following acceptance-rejection rule

$$\theta_l^{t+1} = \begin{cases} \theta_l^* & \text{with prob } \min\left\{\frac{f(\theta_l^*)}{f(\theta_l^t)} \; \frac{g(\theta_l^t | \theta_l^*)}{g(\theta_l^* | \theta_l^t)}, 1\right\} \\ \theta_l^t & else \end{cases}$$

With

$$\frac{f(\theta_l^*)}{f(\theta_l^t)} \; \frac{g(\theta_l^t \mid \theta_l^*)}{g(\theta_l^* \mid \theta_l^t)} = \frac{{\theta_l^*}^{\left(-\frac{N}{\theta_l^*} - a - 1\right)}}{{\theta_l^t}^{\left(-\frac{N}{\theta_l^t} - a - 1\right)}} \; \left[\frac{\Gamma\left(\frac{1}{\theta_l^t}\right)}{\Gamma\left(\frac{1}{\theta_l^*}\right)}\right]^N$$

$$\times \frac{\left(\prod_{n=1}^{N} \frac{1}{s_{nl}}\right)^{\frac{1}{\theta_l^*}} \left(\prod_{n=1}^{N} y_{nl}\right)^{\frac{1}{\theta_l^*}-1}}{\left(\prod_{n=1}^{N} \frac{1}{s_{nl}}\right)^{\frac{1}{\theta_l^t}} \left(\prod_{n=1}^{N} y_{nl}\right)^{\frac{1}{\theta_l^t}-1}} \left[\frac{1-\Phi\left(\frac{-\theta_l^t}{\epsilon_1}\right)}{1-\Phi\left(\frac{-\theta_l^*}{\epsilon_1}\right)}\right]$$

$$\times \exp\left[\frac{-\left(\sum_{n=1}^{N} \frac{y_{nl}}{s_{nl}}\right)-b}{\theta_l^*} + \frac{\left(\sum_{n=1}^{N} \frac{y_{nl}}{s_{nl}}\right)+b}{\theta_l^t}\right]$$

8.3. Sampling the parameter μ_k

We obtain the following conditional law

$$f\left(\mu_k \mid s_{nl}, \sigma_k^2, \mu_0, \sigma_0\right) \propto \exp\left[\frac{(\mu_k-\mu_0)^2}{2\,\sigma_0^2}\right]$$

$$\times \frac{\exp\left[\frac{-\sum_{n=1}^{N}\sum_{l=1}^{L_k}(s_{nl}-\mu_k)^2}{2\,\sigma_k^2}\right]}{\left(1-\Phi\left(-\frac{\mu_k}{\sigma_k}\right)\right)^{NL_k}} \tag{22}$$

Using the following proposal law (Gaussian [31])

$$g\left(x \mid \mu_k^t\right) \propto \frac{1}{\sqrt{2\,\pi}\,\epsilon_3} \exp\left[-\frac{(x-\mu_k^t)^2}{2\,\epsilon_3^2}\right].$$

We obtain the following acceptance-rejection rule

$$\mu_k^{t+1} = \begin{cases} \mu_k^* \text{ with prob } \min\left\{\frac{f(\mu_k^*)}{f(\mu_k^t)} \; \frac{g(\mu_k^t \mid \mu_k^*)}{g(\mu_k^* \mid \mu_k^t)}, 1\right\} \\ \mu_k^t \quad else \end{cases}$$

with

$$\frac{f(\mu_k^*)}{f(\mu_k^t)} \; \frac{g(\mu_k^t \mid \mu_k^*)}{g(\mu_k^* \mid \mu_k^t)} = \left(\frac{1-\Phi\left(-\frac{\mu_k^t}{\sigma_k}\right)}{1-\Phi\left(-\frac{\mu_k^*}{\sigma_k}\right)}\right)^{NL_k}$$

$$\times \exp\left[\frac{-\sum_{n=1}^{N}\sum_{l=1}^{L_k}(s_{nl}-\mu_k^*)^2}{2\,\sigma_k^2} - \frac{(\mu_k^*-\mu_0)^2}{2\,\sigma_0^2}\right]$$

$$\times \exp\left[\frac{\sum_{n=1}^{N}\sum_{l=1}^{L_k}(s_{nl}-\mu_k^t)^2}{2\,\sigma_k^2} + \frac{(\mu_k^t-\mu_0)^2}{2\,\sigma_0^2}\right]$$

8.4. Sampling the parameter σ_k^2

We obtain the following conditional law

$$f\left(\sigma_k^2 \mid s_{nl}, \mu_k, \alpha_0, \beta_0\right) \propto \frac{\left(\frac{1}{\sigma_k^2}\right)^{\frac{NL_k}{2}+\alpha_0+1}}{\left(1-\Phi\left(-\frac{\mu_k}{\sigma_k}\right)\right)^{NL_k}}$$

$$\times \exp\left[\frac{-\sum_{n=1}^{N}\sum_{l=1}^{L_k}(s_{nl}-\mu_k)^2}{2\,\sigma_k^2} - \frac{\beta_0}{\sigma_k^2}\right] \tag{23}$$

Using the following proposal law (positive truncated Gaussian [30])

$$g\left(x \mid (\sigma_k^2)^t\right) \propto \frac{\frac{1}{\sqrt{2\,\pi}\,\epsilon_4} \exp\left[-\frac{(x-(\sigma_k^2)^t)^2}{2\,\epsilon_4^2}\right]}{1-\Phi\left(\frac{-(\sigma_k^2)^t}{\epsilon_4}\right)}.$$

We obtain the following acceptance-rejection rule

$$(\sigma_k^2)^{t+1} = \begin{cases} \sigma_k^{2*} \text{ with prob } \min\left\{\frac{f(\sigma_k^{2*})}{f((\sigma_k^2)^t)} \; \frac{g((\sigma_k^2)^t \mid \sigma_k^{2*})}{g(\sigma_k^{2*} \mid (\sigma_k^2)^t)}, 1\right\} \\ (\sigma_k^2)^t \quad else \end{cases}$$

with

$$\frac{f(\sigma_k^{2*})}{f((\sigma_k^2)^t)} \; \frac{g((\sigma_k^2)^t \mid \sigma_k^{2*})}{g(\sigma_k^{2*} \mid (\sigma_k^2)^t)} = \left(\frac{1-\Phi\left(-\frac{\mu_k}{\sigma_k^t}\right)}{1-\Phi\left(-\frac{\mu_k}{\sqrt{\sigma_k^{2*}}}\right)}\right)^{NL_k}$$

$$\times \left(\frac{(\sigma_k^2)^t}{\sigma_k^{2*}}\right)^{\frac{NL_k}{2}+\alpha_0+1} \exp\left[\frac{-\sum_{n=1}^{N}\sum_{l=1}^{L_k}(s_{nl}-\mu_k)^2}{2\,\sigma_k^{2*}} - \frac{\beta_0}{\sigma_k^{2*}}\right]$$

$$\times \exp\left[\frac{\sum_{n=1}^{N}\sum_{l=1}^{L_k}(s_{nl}-\mu_k)^2}{2\,(\sigma_k^2)^t} + \frac{\beta_0}{(\sigma_k^2)^t}\right] \left[\frac{1-\Phi\left(\frac{-(\sigma_k^2)^t}{\epsilon_3}\right)}{1-\Phi\left(\frac{-\sigma_k^{2*}}{\epsilon_3}\right)}\right]$$

8.5. Sampling the parameter z_l

We have

$$P\left(z_l = k \mid s_{nl}, \sigma_k^2, \mu_k\right) \propto \frac{\frac{1}{(\sqrt{2\,\pi}\,\sigma_k)^N} \exp\left[-\sum_{n=1}^{N}\frac{(s_{nl}-\mu_k)^2}{2\,\sigma_k^2}\right]}{\left[1-\Phi\left(\frac{-\mu_k}{\sigma_k}\right)\right]^N} \tag{24}$$

We must first calculate the proportions associated with each value of k using 24, and associate them intervals whose length is proportional to these values in the interval [0,1].

We generate after a uniform variable as follows

$$u \sim \mathrm{U}[0,1]$$

And finally, z_l is updated by the value of k associated with the interval selected by u.

9. REFERENCES

[1] T. D. Menge, B. P. Hibler, M. A. Cordova, K. S. Nehal, and A. M. Rossi, "Concordance of handheld reflectance confocal microscopy (rcm) with histopathology in the diagnosis of lentigo maligna (lm): A prospective study," *Journal of the American Academy of Dermatology*, vol. 74, no. 6, pp. 1114–1120, 2016.

[2] R. Hofmann-Wellenhof, E. M. Wurm, V. Ahlgrimm-Siess, S. K. E. Richtig, J. Smolle, and A. Gerger, "Reflectance confocal microscopy-state-of-art and research overview," *Seminars in Cutaneous Medecine and Surgery, Elsevier*, vol. 28, pp. 172–179, 2009.

[3] A. Halimi. "Modélisation et traitement statistique d'images de microscopie confocale: application en dermatologie ," (Doctoral dissertation), 2017

[4] A. Halimi, G. J. H. Batatia, J L.Digabel, and J.-Y. Tourneret, "Technical report associated with the paper "An unsupervised Bayesian approach for joint reconstruction and classification of cutaneous reflectance confocal microscopy images"," University of Toulouse, France, Tech. Rep., Feb 2017.

[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," *British journal of Dermatology.*, vol. 170, pp. 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.*, vol. 71, pp. 49–55, 2014.

[7] P. Guitera, L. Haydu, S. Menzies, and al, "Surveillance for treatment failure of lentigo maligna with dermoscopy and in vivo confocal microscopy: new descriptors." *British journal of Dermatology.*, vol. 170, pp. 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," *Dermatologic Surgery*, vol. 40, no. 3, pp. 247–256, 2014.

[9] B. Hibler, M. Cordova, R. Wong, and A. R. ., "Intraoperative real-time reflectance confocal microscopy for guiding surgical margins of lentigo maligna melanoma," *Dermatol Surg.*, vol. 41, pp. 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*, vol. 14, no. 9, pp. 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, pp. 79 041A–79 041A.

[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, pp. 820 702–820 702.

[13] 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, pp. 1–8.

[14] 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*, vol. 11, no. 4, p. e0153208, 2016.

[15] 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, pp. 968 908–968 908.

[16] 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*, vol. 25, no. 5, pp. 554–558, 2011.

[17] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings algorithm," *The american statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[18] W. R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov chain Monte Carlo in practice*. CRC press, 1995.

[19] D. Fink, "A compendium of conjugate priors," *See http://www. people. cornell. edu/pages/df36/CONJINTRnew% 20TEX. pdf*, p. 46, 1997.

[20] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, July 2008.

[21] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

[22] C. P. Robert and D. Cellier, "Convergence control of MCMC algorithms," in *Discretization and MCMC Convergence Assessment*, C. P. Robert, Ed. New York: Springer Verlag, 1998, pp. 27–46.

[23] A. Gelman and D. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, no. 4, pp. 457–511, 1992.

[24] S. J. Godsill and P. J. W. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech, Audio Processing*, vol. 6, no. 4, pp. 352–372, 1998.

[25] P. M. Djurić and J.-H. Chun, "An MCMC sampling approach to estimation of nonstationary hidden markov models," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1113–1123, 2002.

[26] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. London: Chapman & Hall, 1995.

[27] M. Wiltgen, A. Gerger, C. Wagner, J. Smolle, *et al.*, "Automatic identification of diagnostic significant regions in confocal laser scanning microscopy of melanocytic skin tumors," *Methods of Information in Medicine*, vol. 47, no. 1, pp. 14–25, 2008.

[28] C. P. Robert, "Simulation of truncated normal variables," *Statistics and computing*, vol. 5, no. 2, pp. 121–125, 1995.

[29] G. O. Roberts, J. S. Rosenthal, *et al.*, "Optimal scaling for various Metropolis-Hastings algorithms," *Statistical science*, vol. 16, no. 4, pp. 351–367, 2001.

[30] L. Chaari, J.-Y. Toumeret, and C. Chaux, "Sparse signal recovery using a bernoulli generalized gaussian prior," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 1711–1715.

[31] L. Chaari, J.-Y. Tourneret, C. Chaux, and H. Batatia, "A hamiltonian monte carlo method for non-smooth energy sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5585–5594, 2016.