Unsupervised numerical characterisation in determining the borders of malignant skin tumours from spectral imagery

Ilkka Pölönen,Tero Tuovinen, Hannu-Heikki Puupponen, Mari Salmivuori, Mari Grönroos, and Noora Neittaanmäki

Abstract For accurate removal of malignant skin tumors, it is crucial to assure the complete removal of the lesions. In case of certain ill-defined tumors, it is clinically challenging to see the true borders of the tumor. In this paper, we introduce several computationally efficient approaches based on spectral imaging to guide clinicians in delineating tumor borders. First, we present algorithms that can be used effectively with simulated skin reflectance data. By using simulated data, we gain detailed information about the sensitivity of the different approaches and how variables defined by algorithms act in the skin model. Secondly, we demonstrate the performance of the algorithms with spectral images taken *in-vivo* and representing two types of skin cancers with ill-defined borders, namely lentigo maligna and aggressive basal cell carcinoma. The results can be used as a guideline for developing software for the fast delineation of skin cancers.

1 Introduction

Incidence of skin cancer is on a rise, leading to a significant healthcare burden and costs, not only in countries with high levels of ultraviolet (UV) exposure, but also in northern countries, where the light skin phenotype, which is one of the risk factors for skin cancer, is common [7, 16, 3]. Skin cancers are divided to cutaneous melanoma and non-melanoma skin cancers (NMSC), which includes squamous cell

Ilkka Pölönen, The Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland, e-mail: ilkka.polonen@jyu.fi · Tero Tuovinen, The Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland, e-mail: tero.tuovinen@jyu.fi · Hannu-Heikki Puupponen, Airbus DS Finland Oy, Jyväskylä, Finland · Mari Salmivuori, Joint Authority for Päijät-Häme Social and Health Care, Dermatology and allergology, Lahti, Finland · Mari Grönroos, Joint Authority for Päijät-Häme Social and Health Care, Dermatology and allergology, Lahti, Finland · Noora Neittaanmäki, Department of Pathology, Institute of Biomedicine and Department of Dermatology and Venereology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

carcinoma (SCC) and basal cell carcinoma (BCC). NMSC are the most common cancer in the fair-skinned population, and approx. 80 % of NMSCs are BCCs [19]. Currently, between two to three million NMSC and 132,000 melanoma cases are reported annually (WHO webpage) [28].

SCC and BCC are divided into low and high-risk tumors on the basis of histopathology, anatomic location, and the size of the tumor [27, 26]. BCCs have indolent and aggressive histological growth patterns [27]. Cutaneous melanoma has four main subtypes, which are superficial, nodular, lentigo maligna melanoma (LMM) and acral lentiginous melanoma, and the precursor of LMM is lentigo maligna (LM), i.e., *in-situ* melanoma [9]. Out of these classes of tumors, especially aggressive BCCs with ill-defined borders, and LM/LMM with areas of subtle or lacking pigmentation, can have subclinical extensions, and thus these tumors pose a clinical problem in defining the tumor borders.

Currently skin cancer diagnosis is based on histopathological examination of skin biopsies obtained under clinical suspicion. Melanomas, SCCs and high-risk BCCs are treated with a surgical excision [27, 26, 9]. Ill-defined skin tumors may require several re-excisions before complete histological removal with sufficient margins is achieved. With ill-defined borders and possible subclinical extensions, there is a risk of recurrence due to incomplete removal of the tumor [6]. Micrographic surgery is developed to achieve better treatment results in ill-defined tumors or tumors in challenging locations, by examining 100 % of the tumor margins, but this technique demand resources and a trained team of specialists [27, 9, 26]. With non-invasive imaging technology, such as spectral imaging, it may be possible to delineate preoperative tumor border assessment might be more accurate, avoiding resource-consuming re-excisions, or the need for a special operation technique.

The use of spectral imaging for skin cancer analysis, in visible light and in near infrared, is currently under active research. The origin of spectral imaging is in the field of remote sensing [5]. The technology has been applied to different fields of science, mainly in applications where there is a need to detect and recognize objects and outliers from images [8]. Spectral imaging can be used, e.g., to estimate surface moisture of peat production [11], to plan a fertilization task for farm machinery [15], and to authenticate art or to conduct crime scene studies [18].

In medical use, skin cancer analysis is one of the most promising applications for hyperspectral imaging. It has been used succesfully to delineate skin tumors using supervised or semi-supervised approaches [22, 29]. It also has been applied to distinguish between *in-situ* and malignant melanoma [21].

In this paper, we explore several approaches for numerically characterising skin tumor borders to help clinicians detect subclinical parts of the tumor. Our approach works in a fully unsupervised manner, by using a novel spectral imaging system. We demonstrate our method and imaging system by collecting spectral data *in-vivo* from a tumor in a clinical setup. Section 2 presents how the optical spectra of the skin can be simulated. Section 3 describes how spectral images are acquired. The section 4 describes the used and developed algorithms in detail. Section 5 illustrates the results with simulated and imaged spectral data. In the results, the hyperspectrally

defined tumour borders are compared to clinically assessed ones, and results were confirmed by histopathology. Finally, sections 6 and 7 present the conclusions.

The study protocol was approved by the local ethics committee (TAYS/R14120) and followed the Declaration of Helsinki. Volunteering patients were informed orally, and they provided their written informed consent. Two demostrative patient cases were imaged and excised at the department of Dermatology and Allergology in Joint Authority for Päijät-Häme Social and Health Care Group, Lahti, Finland.

2 Mathematical model of the human skin

Skin consists of several layers. The two major layers are called dermis and epidermis. This layered structure can be modelled based on the Kubelka–Munk model for light propagation [17], which is based on the relation between scattering and the absorption coefficient of the layers as well as overall reflectance. The Kubelka–Munk theory describes radiation in diffuse scattering media via an energy transport equation [14]. Based on this model, it is possible to make quantitative studies about absorption, scattering and luminescence in diffuse scattering media, as illustrated in Figure 1. A similar approach has been used previously in the article by [14]. The optical properties of skin have been studied widely in traditional spectroscopy research (see e.g. articles by [2, 13]). There the focus is on the light propagation model, which consists of the structural properties of several layers, where the most important layers are epidermis, dermis and subcutaneous fat [1]. Basically it is possible to represent each layer as as a combination of absorption and scattering properties of different chromophores in the skin.

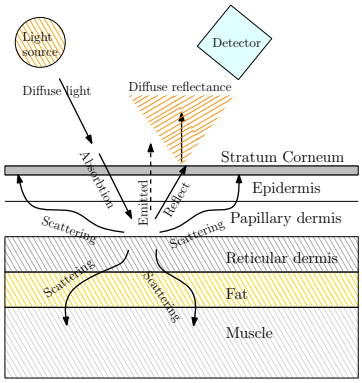

Fig. 1 Schematic illustration of the behaviour of light encountering human skin.

There is a wide range of potential candidates for the model (see e.g. [14]), but in this study, one of the simplest and a numerically inexpensive approach has been chosen. The major chromophore of epidermis is melanin [23], and in dermis, oxygenated and de-oxygenated haemoglobin.

Table 1 The seven parameters and their operating range based on literature [14].

Skin choromophores and parameters used in the equations	**Symbol**
Hemoglobin oxygen saturation	S
Avarage blood volume fraction	B
Water content	W
Bilirubin concentration	C_{bili}
β-carotene concentration	C_β
Fat content	F
Melanosome volume fraction	M
Epidermis thickness	d_{epi}
Dermis thickness	d_{dermis}

According to [13], the absorption μ_a of each skin layer can be characterized as a linear mixture of different chromophores as follows:

$$\begin{aligned}\mu_a =& BS\mu_{a,oxy} + B(1-S)\mu_{a,deoxy} + W\mu_{a,water} + \\ & F\mu_{a,fat} + M\mu_{a,mel} + 2.3(C_{bili}\mu_{a,bili} + C_\beta\mu_{a,\beta}),\end{aligned} \quad [\mathrm{cm}^{-1}] \quad (1)$$

where $\mu_{a,\_}$ are known absorption coefficients of chromophores. Other parameters are listed in Table 1. [1].

The absorption model is simplified. Only the mixture of main chromophores and a baseline spectrum have been used:

$$\begin{aligned}\mu_a =& BS\mu_{a,oxy} + B(1-S)\mu_{a,deoxy} \\ & + M\mu_{a,mel} + (1-M-B)\mu_{a,base},\end{aligned} \quad [\mathrm{cm}^{-1}] \quad (2)$$

where the melanin absorption coefficient is

$$\mu_{a,mel} = 6.6 \times 10^{11} \lambda^{-3.33} \quad [\mathrm{cm}^{-1}] \quad (3)$$

and the baseline absorption coefficient is

$$\mu_{a,base} = 0.244 + 85.3 \exp^{\frac{-(\lambda-164)}{66.2}} . \quad [\mathrm{cm}^{-1}] \quad (4)$$

Both estimates were originally presented by [12]. When considering healthy skin as a two-layered structure, it follows that in the epidermal layer $B = 0$, and in the dermal layer $M = 0$. Cancerous tissue complicates the structure. For example, in advanced

[1] Haemoglobin absorption coefficients used for this study were downloaded from http://omlc.org/spectra/index.html.

malignant melanoma, blood may be found in the epidermal layer, and melanocytes in the dermal layer.

Skin scattering can be approximated using Mie theory as follows:

$$\mu_s' = a\left(\frac{\lambda}{500(nm)}\right)^{-b}, \qquad [\mathrm{cm}^{-1}] \quad (5)$$

where μ_s' is the reduced scattering, a is the scaling factor and b the scattering power. Variance between Mie and Rayleigh scattering is dependent on the diameter of fibres in the skin. If the majority of skin fibres are very small, then the fraction of Rayleigh scattering f_{Ray} is large. For example, low f_{Ray} could indicate that the amount of collagen is high in a certain spot of skin. The skin anisotropy factor g varies between $0.5 and 0.9$. The scattering for each skin layer is $\mu_s = (1-g)\mu_s'$.

The equation for reflectance R for one layer can be represented as a function of layer thickness d_{layer}, absorption μ_a and scattering μ_s [14]:

$$R_{layer}(\lambda) = \frac{(1-\beta)^2)(\exp^{K_{layer}d_{layer}} - \exp^{-K_{layer}d_{layer}})}{(1+\beta)^2\exp^{-K_{layer}d_{layer}} - (1-\beta)^2\exp^{-K_{layer}d_{layer}}} \quad (6)$$

Transmittance T can be represented as

$$T_{layer}(\lambda) = \frac{4\beta}{(1+\beta)^2\exp^{K_{layer}d_{layer}} - (1-\beta)^2\exp^{-K_{layer}d_{layer}}} \quad (7)$$

where K is the backward flux

$$K_{layer} = \sqrt{k_{layer}\left(k_{layer} + 2\times s_{layer}\right)} \qquad [\mathrm{cm}^{-1}] \quad (8)$$

and β is the forward flux

$$\beta_{layer} = \sqrt{\frac{k_{layer}}{k_{layer} + 2s_{layer}}}, \quad (9)$$

and

$$k_{layer} = 2\times\mu_{a_{layer}}, \qquad s_{layer} = 2\times\mu_{s_{layer}}. \qquad [\mathrm{cm}^{-1}] \quad (10)$$

These equations follow from the Kubelka–Munk theory.

The total reflection can be expressed as

$$R_{total} = R_{epidermal}R_{dermal} = R_{epidermal} + \frac{T_{dermal}^2 R_{epidermal}}{1 - R_{epidermal}R_{dermal}}. \quad (11)$$

This model will be later used to simulate skin to evaluate section 4 algorithms performance, when there are small changes in the model's parameter.

3 Spectral imaging

The imaging sensor captures the spectrum of reflecting substances, and creates a datacube spanning both spatial and spectral axes. , which can be analyzed further. Each spatial pixel in the image forms a spectrum in a spectral dimension, as illustrated in Figure 2. The basic principle of light spectroscopy is that substances reflect different wavebands of light differently. Therefore many substances have a unique spectral signature. In spectral analysis, the goal is to recognize different substances based on how they absorb and scatter light. Hyperspectral imaging diverges from normal reflectance spectroscopy in that it produces a composite image of all the spectra.

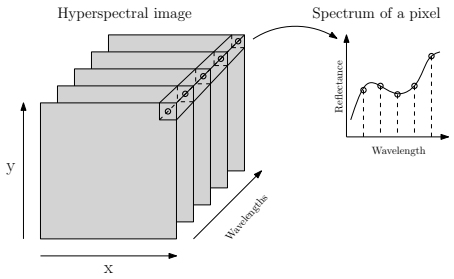

Fig. 2 Illustration of a hyperspectral image. Each pixel in the image forms a spectrum through the imaged wavebands.

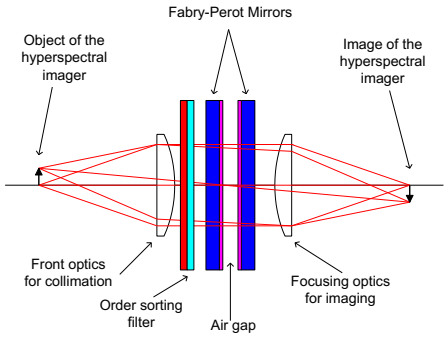

Fig. 3 Illustration of the Fabry–Perot interferometer structure inside the imager.

The instrument used in this study was a small and lightweight hyperspectral imager developed by VTT Technical Research Centre of Finland. Spectral separation in the device is carried out through a piezo-actuated Fabry–Perot interferometer (FPI). The component is hermetically sealed in a metal can filled with nitrogen. Both the parallelism and the distance between the mirrors of the Fabry–Perot interferometer

are controlled with high accuracy. This is achieved with three closed-loop control channels at the edges of the mirror plates. Each channel has a piezoelectric actuator with a closely positioned capacitive measurement point, which is used to determine the mirror separation. Each channel is controlled with nanometer accuracy to obtain the desired parallelism and air gap between the mirrors. The structure of the interferometer system is illustrated in Figure 3, and the optical concept is described in Figure 4.

The spectral imager uses a CMOS sensor. Combined with the piezo-actuated Fabry–Perot interferometer, this enables quite a high speed of image acquisition. Depending on the lighting conditions, the exposure time for each plane can vary from 0.1 to 1,000ms. In this setup, we used an exposure time of 30ms. The whole spectral cube was recorded in approximately one second. This depends on how many wavebands are to be included in the measurement, and the target resolution. Each band's center wavelengths and full width at half of the maximum (FWHM) can be found in Table 3 in the Appendix. FWHM varies between 11 to 45 nm. Currently we have not compensated for or taken into account overlap between wavebands.

Other technical specifications can be found in Table 2 in the Appendix. For the purposes of this study, four times four pixel binning was in place, giving a final image plane resolution of 320 by 240 pixels.

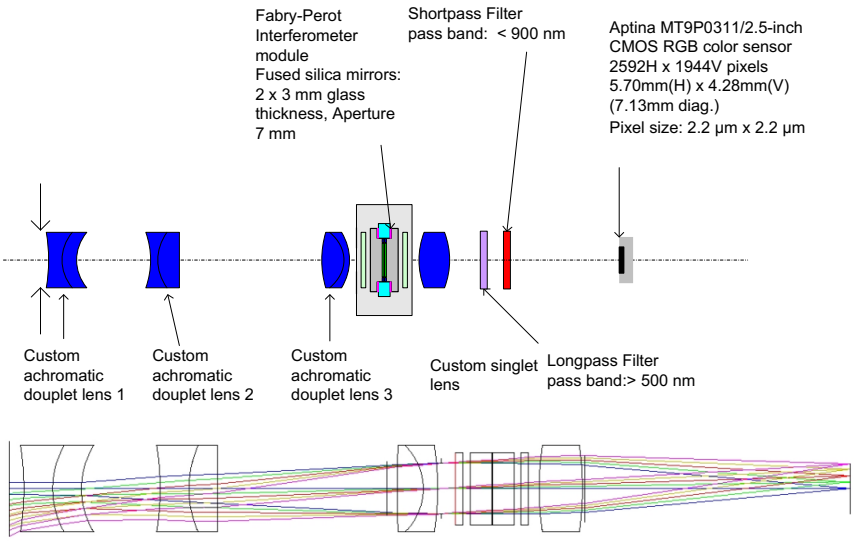

Fig. 4 The optical concept of the hyperspectral imager.

The camera was placed on a custom made holder, which also integrated a fiber-optic ringlight around the camera's lens system. The ringlight was attached to a halogen-based fiber-optic illuminator (Dolan-Jenner, Fibre-light DC950). The illuminator had an infrared filter, which was removed to improve signal efficiency in the

infrared region. A diffuse filter sheet was placed against the face of the ringlight to achieve a more consistent light distribution pattern and to reduce specular reflections from the skin surface. To exclude external light sources that could interfere with the measurements, and to maintain a constant, reproducible distance between the subject and the imager, an interchangeable cone threaded onto the handle was utilized.

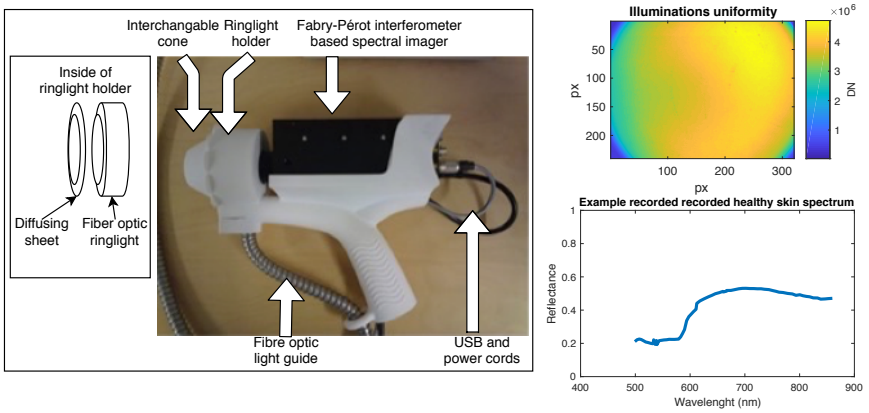

Fig. 5 *Left*: Overview of the spectral imager system. *Top right*: Illumination uniformity map of a white reference target, integrated over wavebands. *Bottom right*: An example random healthy skin spectrum.

The captured intensity I is converted into reflectance $\mathbf{X}$ by dividing the recorded data with a white reference target for all wavelengths λ , $\mathbf{X}_\lambda = (I_\lambda - I_{dark})/(I_{\lambda,white} - I_{dark})$, where I_λ is the recorded raw digital image, $I_{\lambda,white}$ is the recorded white reference target, and I_{dark} is the dark current of the imaging sensor. The exposure of the hyperspectral image is adjusted with the white reference. In demonstrations, we have cropped the image border so that only the region of interest is included in further analysis.

4 Computational methods

This section presents the algorithms for the analysis. Let $\mathbf{x} = [x_1 x_2 \dots x_d] \in [0, 1]$ be a recorded reflectance spectrum, where d is the number of wavebands. Moreover, let $\mathbf{X} = (\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_n)^\mathsf{T}$ be a spectral image, with the pixels linearly indexed and n the total number of pixels in the spectral image.

4.1 Modified standard normal variate correction algorithm

Originally the standard normal variate algorithm [4] was invented to reduce spectral noise and eliminate the effect of the background. The algorithm was designed for single point spectrometers, where samples are measured as a time series. In the imaging sense, this is not very relevant, because there are better ways to standardise data points. However, if there is a situation where a quite large area of an image is spectrally relatively homogenous, the algorithm can be used to characterise differences in the image, without prior information about the spectra of the homogenous background.

First, we calculate the mean in the spectral direction for each pixel in the image as

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i. \tag{12}$$

Next the standard deviation is calculated for each pixel,

$$x_i^{STD} = \sqrt{\frac{\sum_{1=j}^{d}(x_j - \overline{x}_j)^2}{d}}. \tag{13}$$

Now, we transform the centers $\mathbf{SNV}_{ij}$ for each spectra in the image are calculated, so that

$$\mathbf{SNV}_{ij} = \frac{\mathbf{x}_i - \overline{\mathbf{x}}}{x_i^{STD}}. \tag{14}$$

Finally, we sum over all wavebands

$$MSNV_i = \sum_{j=1}^{d}\mathbf{SNV}_{ij} \tag{15}$$

The $MSNV_i$ values are calculated for the whole image area, and the results are reshaped to the size of the original images. This image is normalized to the range $[0, 1]$.

4.2 Linear unmixing

In this method we assume that the measured spectra are a linear mixture of some spectra, which are characteristic to the image and present in it. Let

$$\mathbf{X} = \mathbf{MY}, \tag{16}$$

where $\mathbf{M} \subset \mathbb{R}^{k\times n}$ is the mixing matrix, and $\mathbf{Y} \subset \mathbb{R}^{k\times d}$ are the characteristic spectra, which are called endmembers. To solve M from the equation (16), it is necessary to determine the endmembers Y.

Algorithm 1 Iterative Vertex Component Analysis (IVCA)

Input: Spectral data **X**
Output: endmembers, abundance maps

```
Calculate PCA or SVD.
for k = 2 to 6
        Fix first endmember
        for i = 2 to k
                Do affine transformation.
                Fix i:th endmember.
        end
        Save endmembers.
        Calculate abundance maps using FVA.
        Reconstruct the data multiplying endmembers with abundances.
        Calculate the reconstruction error using the Frobenius norm.
end
Find the smallest construction error and select the endmembers based on this
```

Vertex component analysis (VCA) [20] is widely used in the endmember extraction method. It can be considered a state-of-the-art algorithm. It is based on the geometry of spectral data. The main assumption behind VCA is that the endmembers are vertices of a convex hull, which covers all spectra. The algorithm assumes that there are pure endmembers in the image. The original VCA needs to know how many endmembers there are in the data set. Thus, the method itself is not fully automatic. However, VCA is a relatively fast algorithm. The most computationally expensive part is to either calculate a singular value decomposition (SVD), or perform a principal component analysis (PCA), which reduces the dimensionality of the data. After this phase, the algorithm fixes the maximum value of the first principal component as the first endmember and performs affine transformation with it. The second endmember is then the maximum of this transformed data. This continues until all k endmembers are found.

After the endmember extraction is done, we need to solve the mixing matrix **M**, which describes the proportion of endmembers in each pixel. Linear equations can be solved in multiple different ways. To execute inversion computationally in an efficient manner, we may apply the Filter Vector Algorithm (FVA) [24]. FVA is derived from the least squares method. Basically, from the found endmembers **Y**, the first task is to calculate a filter matrix using the pseudoinverse

$$\mathbf{F} = (\mathbf{Y}^t\mathbf{Y})^{-1}\mathbf{Y}^t. \tag{17}$$

Now, the mixing matrix is simply

$$\mathbf{M} = \mathbf{F}\mathbf{X}. \tag{18}$$

Because the filter vectors are an orthogonal projection to the subspace formed by the endmembers, it holds that

$$\mathbf{f_i y_j} = \begin{cases} 1, i = j \\ 0, i \neq j. \end{cases} \tag{19}$$

and for a single spectrum $\mathbf{x_j}$,

$$\mathbf{f_i x_j} = \mathbf{f_j} \left(\sum_{i=1}^{k} m_i \mathbf{y_i} + n \right) = \sum_{i=1}^{k} f_\mathbf{j} \mathbf{y_i} m_i + f_j n = \mathbf{m_j}, \tag{20}$$

where n is noise and $\mathbf{m_j}$ is the proportion of endmembers $\mathbf{Y}$ in pixel j. These proportions (M) are reshaped as images, which are termed abundance maps of the corresponding endmember.

Because of the assumption that the data is linearly mixed, it is possible to reconstruct data using the found endmembers. The reconstructed data is

$$\overline{\mathbf{X}} = \mathbf{MY}. \tag{21}$$

To automate VCA, it is necessary to approximate how many endmembers there are. One option is to minimize the reconstruction error between the original recorded and reconstructed data sets, subject to the number of endmembers k, using the Frobenius norm, i.e.

$$\min_k ||\mathbf{X} - \overline{\mathbf{X}}||_{FRO}, \tag{22}$$

where k varies between 2 and 6. Based on experience, VCA usually detects two to four endmembers from skin with a benign or malignant tumor. In the numerical examples, which are processed using Algorithm 1, the dimension reduction has been applied to full data. For the iterations, it is necessary to have as many dimensions as the largest used k value will be. Then the procedure continues by running VCA in all iterations, and calculating each time the abundance maps and the reconstruction error. The reconstruction error is the smallest when the number of endmembers correspond to the real number of unique spectra in the image. With the found best fitting endmembers, the abundance maps are calculated.

4.3 Linear unmixing of estimated single scattering albedo

Section 2 shows that the skin reflectance spectra are actually a nonlinear mixture of the absorbance of different chromophores and scattering properties of skin cells. Thus it is a legitimate question whether non-linear models should be applied. Hapke [10] suggests a rapid semiquantitative estimation formula to convert reflectance to singe scattering albedo (SSA),

$$\omega = \frac{4\mathbf{x}}{(1 + \mathbf{x})^2}, \tag{23}$$

where **x** is the diffuse reflectance spectrum. Now, the SSA ω of an intimate mixture is actually a linear mixture of the SSAs of different components, and abundances are proportions of mass fractions of some materials.

It is possible to reduce the non-linearity of reflectance data using the equation (23). After the reduction, all that is needed is to run IVCA for the data to solve different components of the SSAs and calculate their abundances.

4.4 Closed form chromophore specific approximation for estimated single scattering albedo

By definition, a single scattering albedo consists of scattered and absorbed light, so that

$$\omega(\lambda) = \frac{\mu_s(\lambda)}{\mu_a(\lambda) + \mu_s(\lambda)}, \tag{24}$$

where μ_s is the scattering coefficient and μ_a is the total absorption coefficient.

Algorithm 2 Closed form approximation for the estimated SSA of skin

Input: ω, wavelengths, absorption coefficients for skin chromophores $\mu_{a,i}$
Output: proportion of different chromophores, μ_s, $\widehat{\mu_a}$

1. Calculate initial scattering
2. Calculate first iteration of μ_a (eq. 25).
3. Solve chromophore proportions using FVA.
4. Reconstruct absorption coefficient by multiplying chromophore absorption coefficients by their proportions.
5. Calculate scattering μ_s (eq. 26).
6. Calculate second iteration of μ_a (eq. 25).
7. Solve chromophore proportions using FVA.
8. Reconstruct absorption coefficient again by multiplying chromophore absorption coefficients by their proportions.

It is possible to approximate the albedo of skin using the equation (23). From the equation (24), it is evident that

$$\mu_a(\lambda) = \frac{\mu_s(\lambda)}{\omega(\lambda) - \mu_s(\lambda)} \tag{25}$$

and similarly

$$\mu_s(\lambda) = \frac{\omega(\lambda)\mu_a(\lambda)}{1 - \omega(\lambda)}. \tag{26}$$

As was pointed out in Section 2, total absorption is a linear combination of different chromophores of the skin. Parameters for the skin scattering equation (5) are found from the literature. Based on these literature values, initial scattering μ_s for

skin is calculated. The absorption coefficient for skin is estimated using the equation (25) and the linear mixture of the chromophores is solved from μ_a using FVA. Next, the procedure continues by simulating skin absorption by exploiting the determined mixing matrix $\mathbf{m}$, so that $\widehat{\mu_a} = \sum_i m_i \mu_{a,i}$. A better estimation for μ_s is obtained from the equation (26) with $\widehat{\mu_a}$ and ω. Again, we solve a new mixing matrix m for the absorption coefficients of different chromophores, and simulate the total absorption coefficient of skin.

As a result, an estimation for the proportion of different chromophores, total scattering and absorption coefficients of the skin and the tumor for each pixel in the image are established. From the ratio of oxygenated and deoxygenated haemoglobin, it is possible to calculate an estimate for oxygen saturation.

5 Results

5.1 Simulated spectral cube

The data cube was simulated using the method described in Section 2. We simulated five different spectra, which were used to create a data cube with different kind of areas as shown in Figure 6. First, healthy skin area was simulated so that $g = 0.73$, $S = 0.795$, $B = 0.051$, $M = 0.21$, $a = 0.16$, $b = 0.6$, $d_{dermis} = 0.0011$ and $d_{epidermis} = 0.00091$. Here the parameter g is the skin anisotropy, S is oxygen saturation, B is the blood volume fraction, M is the melanocyte fraction, a is the scattering scaling factor, b is the scattering power, d_{dermis} is the thickness of the dermal layer and $d_{epidermis}$ is the thickness of the epidermal layer.

Four differing areas were created by modifying parameters of healthy skin, denoted as background in Figure 6.a. For Area 1, the scattering power b was changed to 0.7. For Area 2, blood volume fraction B was changed to 0.151. For Area 3, melanocyte fraction M was changed to 0.41. Finally, for Area 4, the parameters were changed as follows: $B = 0.151$, $M = 0.25$, $a = 17$ and $b = 0.7$. Each area and background corresponds to one spectra in Figure 6.b. To make the spectral data cube set more realistic, 10 % gaussian noise was added to each pixel.

These small changes should indicate the sensitivity and accuracy of different methods proposed in section 4 for incremental changes in the spectrum.

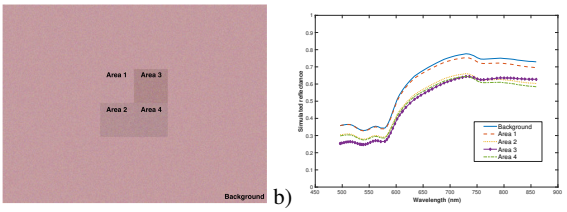

Fig. 6 a) Color image from a simulated spectral cube. In this case all areas except Area 1 are clearly distinguishable from the background. In more realistic cases, these areas have more diffuse and irregular borders. b) Simulated spectra, which correspond to different areas of a sub-image (a) and the parameters used in this area. The spectra have quite similar shape and intensities.

In Figure 6, the spectra given by the reflectance model are presented, including the parameter values used. We observe that the spectra are quite similar to each other. As Figure 6 shows, all areas except Area 1 are clearly visible against the background. In more realistic cases, these areas have usually more diffuse and irregular borders. By looking at Figures 7 and 8, which represent the abundance maps from the linear mixture of the estimated SSA, we may conclude that both methods detect the melanin changes in Area 3. Nevertheless neither of them is able to distinguish changes in scattering (Area 1) or with haemoglobin (Areas 2 and 4).

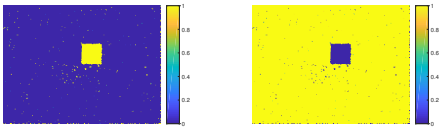

Fig. 7 Abundances of the linear mixture of the detected endmembers of simulated data. *Left*: we notice that Area 3 (with increased melanin) is separated as individual endmembers .*Right*: the background and Areas 1, 2, and 4 are mixed between themselves. Arbitrary units.

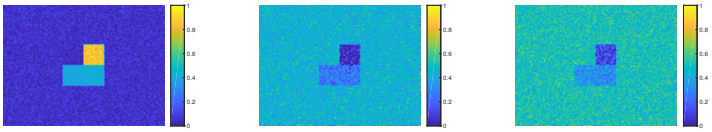

Fig. 8 Abundances of the linear mixtures of the estimated single scattering albedo. *Left*: Area 3 (with higher melanin) has mixed partly with Areas 2 and 4. Higher blood volume in Areas 2 and 4 is visible in the map, but the actual endmember is from Area 3. *Center, Right*: the background and Area 1 are mixed with both endmembers and thus non-separable. Arbitrary units.

Closed form approximation is capable of detecting changes in blood volume and in melanin concentration as illustrated in Figure 9, although it slightly conflates them. The major advantage of the model is that it can detect the small change in

scattering between the background and Area 1. Moreover, the increase of absorption is visible in the mean absorption map.

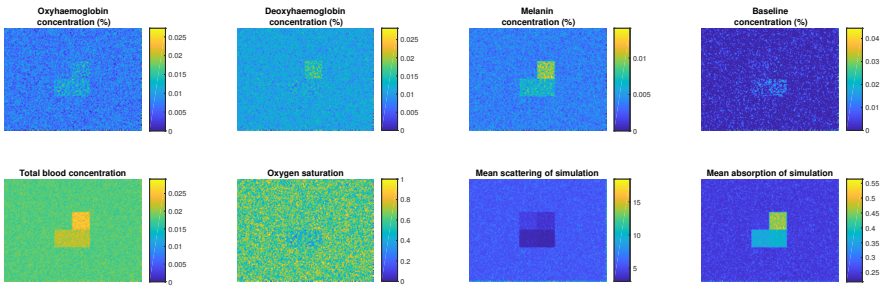

Fig. 9 Chromophore and attribute specific maps for closed form estimation of single scattering albedo. These maps show how chromophores and different scattering attributes have been distributed. For chromophores the range is from 0 to 1, showing their proportion in the tissue. Oxygen saturation is the ratio between Oxy- and de-oxyhaemoglobin. Scattering and absorptions are given in units of mm^{-1}.

The modified SNV algorithm performs well when the issue is detecting changes from the background, as in Figure 10 a). Nevertheless, if the ratio of the background area compared to the target areas decreases, then also the performance of the modified SNV in distinguishing areas decreases, as is revealed by Figure 10 b).

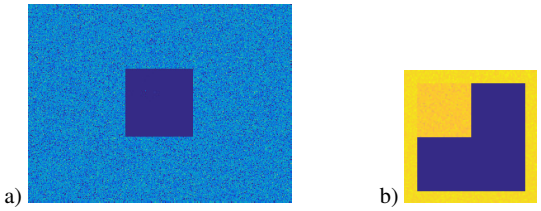

Fig. 10 *a)* The result of the modified SNV algorithm for the whole simulated data cube. The algorithm is capable of distinguishing modified areas from the background. *b)* The result of the modified SNV algorithm for partial simulated data cube. The performance of the algorithm in distinguishing areas is not that good compared to when there is more background in the image. Arbitrary units.

5.2 Basal cell carcinoma

In the clinical part of the study with ill-defined BCCs, the tumour borders were assessed clinically with a dermatoscope (i.e. using a hand-held epiluminescence device). Thereafter tumors were imaged with the hyperspectral imaging system, and

the marginals were excised as a separate 2 mm circumferential strip - both specimens marked with orientation. Histopatholgical analyses were made with traditional methods with vertical sections. This procedure allowed us to confirm histopathologically, if tumor-borders defined by clinician matched with borders assessed from HIS images [25]. So if subclinical extension suspected by HIS was found in the separate marginal strip, this was interpreted as true positive finding, and if tumor-borders were assessed to be smaller with HIS, there was a tumor-free area reported in the histology of the separately excised tumor piece - also interpreted as a true positive finding. In the clinical study reported by Salmivuori et al., there was 12 out of 16 ill-defined BCCs delinated more accurately with the proposed system of hyperspectral imaging compared to the clinical evaluation (four wider and eight smaller by HIS).There were two false positive and two false negative cases.

We present here one case in more detail. In Figure 11, we observe an ill-defined BCC with later confirmation of several aggressive growth patterns in histopathological examination. Black arrows show the direction of the subclinical parts of the lesion. The subclinical extensions were confirmed in histopathology.

Fig. 11 An aggressive basal cell carcinoma. Left: Narrow band RGB image constructed from spectral image. Black arrows show the locations of the subclinical extensions. Right: RGB image of BCC same BCC lesion. Black dot line represent clinical delineation, which was first excised. After that separate 2 mm circumferential strip was removed.

Figure 12 presents the linear mixture model revealing the areas of a subclinical tumor extension. Because melanin is not the main chromophore in BCCs, we assume that the found endmembers are distinguished mostly based on the blood volume fraction in the skin.

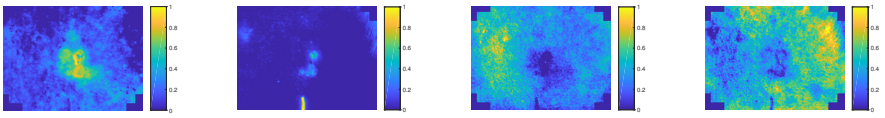

Fig. 12 Abundances of the linear mixture of the detected endmembers of basal cell carcinoma. On the left, the most distinguishing feature seems to correlate with the blood volume fraction. The second image from the left seems to be the marker. The two images on the right seem to divide healthy skin area to two parts. Arbitrary units.

Similar results were also obtained with an estimated single scattering albedo in Figure 13. Here, the algorithm was additionally capable of separating the ulceration area and the marker. Moreover, it is possible to observe differences between the backgrounds, caused by the blood volume fraction.

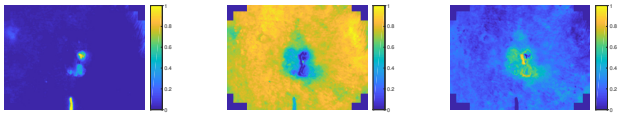

Fig. 13 Abundances of the linear mixtures of estimated single scattering albedo of basal cell carcinoma. *Left*: the ulceration area and the marker are also distinguished as their own endmember. *Center*: the method is able to capture the spectrum of healthy skin. *Right*: the most distinguishing feature seems to correlate with the blood volume fraction. Arbitrary units.

Finally, the chosen chromophore and the attribute-specific closed-form approximation reveal that there is a higher blood volume fraction in the tumor area, also in the subclinical parts. In Figure 14 we observe a clearly visible higher mean absorption and blood volume in the tumor area. The ulcerated area has anomalous behavior.

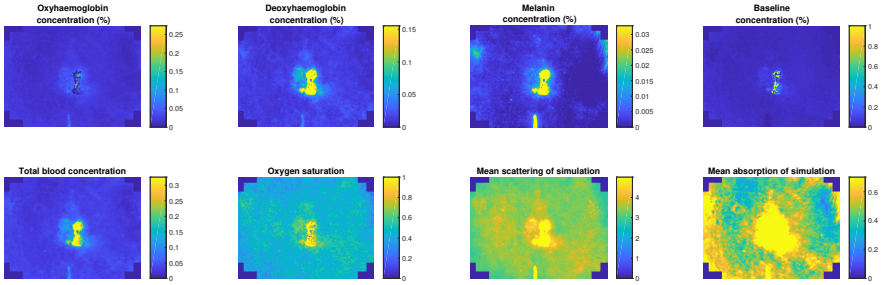

Fig. 14 The chromophore and the attribute-specific closed-form approximation of basal cell carcinoma reveals a higher blood volume fraction in the tumor area, including the subclinical parts. For chromophores, the range is from 0 to 1, showing their proportion in the tissue. Oxygen saturation is the ratio between Oxy- and de-oxyhaemoglobin. Scattering and absorptions are given in units of mm^{-1}.

Modified SNV also separates lesions, as shown in Figure 15, but it also separates some other details as emphasizing that tumors without pigmentation as main chromophore are more difficult to image than pigmented ones..

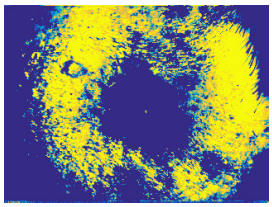

Fig. 15 Modified SNV for basal cell carcinoma.

5.3 *In-situ* melanoma

In the clinical pilot study with LM/LMM, the tumor borders were assessed visually, dermatoscopically and with Wood's light (i.e. a blue light highlighting the pigmentation). The lesions were excised with wide margins. Inking and sutures were used for orientation, and also targeted biopsies of suspected subclinical areas shown by hyperspectral imaging were used. The hyperspectral imaging results were compared to the histopathology. In this clinical part of LM/LMM, the linear mixture model was used [22]. As reported by Neittaanmäki-Perttu et al. delineation by hyperspectral imaging matched the histopathological analysis in 18/19 (94.7 %) of cases, while in 1/19 (5.3 %) cases linear mixture analysis showed lesion extension not confirmed by histopathology (false positives). Compared to clinical examination, HIS defined lesion borders more accurately in 10/19 (52.6 %) of cases. In the rest of the cases the borders were identical. [22]

We present here one case in more detail, also using the other three numerical models, which were developed later [25]. Figure 16 represents a clinical photograph of a lentigo maligna. Black arrows show direction of subclinical parts of the lesion. Subclinical extensions were confirmed by histopathology.

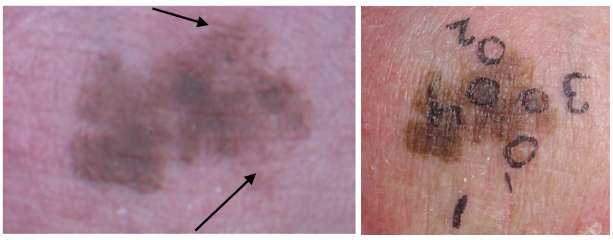

Fig. 16 *In-situ* melanoma: Lentigo maligna. Left: arrow band RGB image constructed from spectral image. Black arrows show the locations of the subclinical extensions. Right: RGB image of same lesions with biopsy locations. Locations 1 and 2 were suspected to be subclinical extensions based on spectral image analysis and later confirmed by histopathology.

The linear mixture model seems to separate melanin and changes in the blood volume fraction to three different endmembers. As Figure 17 shows, it is also capable of characterising visible subclinical changes.

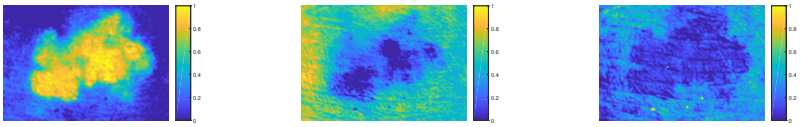

Fig. 17 Abundance maps from the linear mixture model of lentigo maligna. On the left, the procedure is capable of detecting melanin and changes in the blood volume fraction. Moreover, subclinical parts of the lesion are clearly visible. At center and on the right, healthy surrounding skin seems to divide into different endmembers. Arbitrary units.

It was possible to distinguish two endmembers from the estimated single scattering albedo, which seems to correspond to melanin fraction and non-melanin skin. As revealed by Figure 18, subclinical parts are clearly visible.

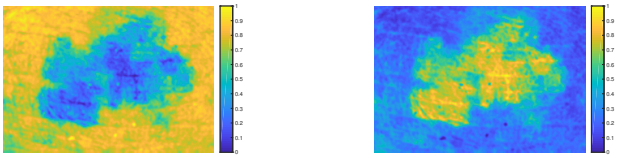

Fig. 18 Abundances of the linear mixtures of estimated single scattering albedo of lentigo maligna. Here separation occurs between normal skin (image on left) and skin containing melanin (image on right). Subclinical parts of the lesion are characterised. Arbitrary units.

The closed-form approximation for a single scattering albedo shows clearly how melanin is distributed in the tumor. In Figure 19, we see difference in the fractions of melanin on subclinical parts, but the real difference seems to be in the scattering at the borders of the lesion. A significant difference can be found between healthy and malicious areas. This could indicate that there are also structural changes in skin cells, and this information could be used for decision-making in the clinical treatment.

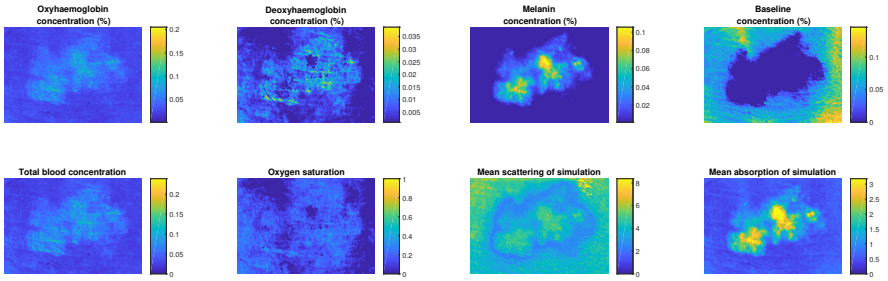

Fig. 19 Closed-form approximation for single scattering albedo for lentigo maligna. Subclinical parts are characterised as melanin and scattering maps. For chromophores the range is from 0 to 1, showing their proportion in the tissue. Oxygen saturation is the ratio between Oxy- and de-oxyhaemoglobin. Scattering and absorptions are given in units of mm^{-1}.

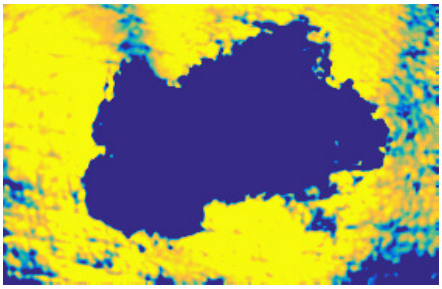

Fig. 20 Modified SNV is capable of detecting subclinical parts of lentigo maligna.

6 Discussion

The proposed algorithms provided reliable results for the characterisation of tumor borders confirmed by histopathology. However, the method still has its limitations. Currently, rounded or irregular surfaces such as the tip of the nose or ears pose challenges to the imaging, while both of them are typical locations for BCC. Imaged spectra from such surfaces are not consistent, but includes interruption from imaging geometry, which also affects the results.

By modifying the imaging setup, it may be possible to reduce different sources of inaccuracies. The current calculation of reflectance is based on a simple ratio between the reflected flux from the skin and the white reference. This increases noise in the data, because there is variation in the optical path length between these surfaces. If the lightning system was modified so that incident light arrives from several directions with known angles, it is possible to calculate the reflectance also using bidirectional reflectance distribution functions. Another potential option is to

estimate single scattering albedo (SSA) using photometric stereo imaging, which provides SSA, surface normal and 3D-surface model as a function of wavelength.

Beside tumor delineation in spatial coordinates, there is also a depth factor, which should be accounted for. Both BCC and melanoma grows also in depth. For example, *in-situ* melanoma develops into invasive malignant melanoma when it has grown through the basal membrane to dermis. Unfortunately, our current fast inversion model does not take the depth factor into account. Adding layers to the model makes it ill-posed and non-linear. This kind of inversion can be solved from the Kubelka–Munk model, but the solution will include some inaccuracies. Within healthy skin, blood and melanocytes appear in their own layers. Thus, the number of degrees of freedom is much smaller than in malignant skin tissue, where melanocytes become mixed into the dermal layer, and blood into the epidermal layer. In theory, it is possible to distinguish the difference between skin layers. For example, if we would see melanocytes in the dermal layer, we could instantly suspect malignant melanoma instead of melanoma *in-situ*. In our clinical pilot study by Neittaanmäki et al. it seems that HIS is capable to distinguish the in situ lentigo maligna from the invasive lentigo maligna melanoma with a positive predictive value of 75%, a negative predictive value of 95%, sensitivity of 90% and specificity of 86.3% [21]. But these are preliminary results and future research is demanded with lager trials.

We emphasize that any numerical model is more or less a compromise that describes the real world, approximated models even more so. The proposed approach approximates heavily both in the model and in the inversion. For example, Hapke's theories have been criticised, but as he cites George Box in [10], *all models are wrong, but some are useful*. Because results from the clinical pilot studies are reasonable [22, 25], the approximations made here can be considered legitimate.

Under these observations, the proposed methods are able to detect clinically ill-defined borders of skin tumors on relatively flat surfaces. We aim to continue our studies further to the direction of developing a system which can handle irregularities on imaging surfaces, and can utilise more precise models with an efficient solution approach for the computational inverse problem.

7 Conclusion

In this paper, several computationally efficient and fast approaches to delineate tumour borders from spectral imagery were presented. We introduced a procedure for the analysis. Algorithms of the procedure were described, and were used effectively with simulated skin reflectance data, and later on, with the analysis and decision making process of real skin cancer cases. By using simulated data, detailed information about sensitivity of the different approaches was gained. We demonstrated the performance of algorithms in two example cases more closely, imaged *in-vivo* from real skin cancers. Results indicate that the proposed algorithms can be used for characterising tumor borders.

Acknowledgements This study was partly funded by the Academy of Finland (grant number 314519).

Appendix

Table 2 Specifications of used Fabry–Perot imager.

Parameter	Imager
Horizontal and vertical FOV (deg.)	$> 36, > 26$
Nominal focal length (mm)	9.3 ± 3 (Custom lenses)
Wavelength range (nm)	$500 - 885$
Spectral resolution at FWHM (nm)	$9 - 40$
Adjustable spectral resolution step	< 1
f-number	< 6.7
Maximum spectral image size (pixels)	2592×1944
Spectral image size with default binning (pixels)	320×240
Camera dimensions (mm)	$62 \times 66 \times 219$
Weight (g)	< 450

Table 3 Used wavelenghts and full width at half of the maximum (FWHM) of each band in nanometers.

Wavelenght	496.2	496.4	499.2	503.2	507.2	510.8	514.6	518.9	523.7	528.4
	532.4	533.5	535.9	538.5	539.6	543.2	547.3	552	557.1	561.8
	566.2	571.5	579.1	582.8	586.1	589.8	593.8	597.7	601.9	606.3
	610.9	612.2	618	625.9	634.1	639.8	646	653.6	659.9	660.8
	666.6	666.9	674.5	682.3	688.9	695.3	701.8	709.1	716.4	723
	729.5	735.6	744	750.5	756.7	762.5	769.8	777.7	786.7	793.4
	800.2	806.2	812.3	818.3	824.9	832.7	838.6	846.5	852.5	859.8
FWHM	32.1	30.3	29.3	45.6	30.2	29.2	30.8	25.2	30.7	20.9
	30.9	31.7	31.2	30.3	30.3	31	30.6	30.7	19.4	29.9
	21.9	30.3	21.4	30.7	21	30.7	21.8	29.4	22.8	29
	23.1	29.4	22	31	20.4	31.9	19.8	31.6	19.8	30.3
	20.2	29.6	21.4	28.5	22.2	28.1	22.4	28.6	21.8	29
	23.3	28.7	24.9	28.3	23	28.6	20.2	27.6	19.2	28.4
	19.7	28.5	19.6	19.5	19.8	20.4	11.2	20	12.9	15.3

References

1. Anderson, R., Hu, J., Parrish, J.: Optical radiation transfer in the human skin and applications in in vivo remittance spectroscopy. In: Bioengineering and the Skin, pp. 253–265. Springer (1981)

2. Anderson, R.R., Parrish, R.R.: The optics of human skin. Journal of Investigative Dermatology **77**, 13–19 (1981). DOI 10.1111/1523-1747.ep12479191
3. Azoury, S.C., Lange, J.R.: Epidemiology, risk factors, prevention, and early detection of melanoma. Surgical Clinics **94**(5), 945–962 (2014)
4. Barnes, R., Dhanoa, M.S., Lister, S.J.: Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied spectroscopy **43**(5), 772–777 (1989)
5. Bioucas-Dias, J.M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., Chanussot, J.: Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and remote sensing magazine **1**(2), 6–36 (2013)
6. Boehringer, A., Adam, P., Schnabl, S., Häfner, H.M., Breuninger, H.: Analysis of incomplete excisions of basal-cell carcinomas after breadloaf microscopy compared with 3d-microscopy: a prospective randomized and blinded study. Journal of cutaneous pathology **42**(8), 542–553 (2015)
7. Cakir, B.Ö., Adamson, P., Cingi, C.: Epidemiology and economic burden of nonmelanoma skin cancer. Facial plastic surgery clinics of North America **20**(4), 419–422 (2012)
8. Chang, C.I.: Hyperspectral data exploitation: theory and applications. John Wiley & Sons (2007)
9. Garbe, C., Peris, K., Hauschild, A., Saiag, P., Middleton, M., Bastholt, L., Grob, J.J., Malvehy, J., Newton-Bishop, J., Stratigos, A.J., et al.: Diagnosis and treatment of melanoma. european consensus-based interdisciplinary guideline–update 2016. European Journal of Cancer **63**, 201–217 (2016)
10. Hapke, B.: Theory of reflectance and emittance spectroscopy. Cambridge university press (2012)
11. Honkavaara, E., Eskelinen, M.A., Pölönen, I., Saari, H., Ojanen, H., Mannila, R., Holmlund, C., Hakala, T., Litkey, P., Rosnell, T., et al.: Remote sensing of 3-d geometry and surface moisture of a peat production area using hyperspectral frame cameras in visible to short-wave infrared spectral ranges onboard a small unmanned airborne vehicle (uav). IEEE Transactions on Geoscience and Remote Sensing **54**(9), 5440–5454 (2016)
12. Jacques, S.L.: Skin optics. Oregon Medical Laser Center News **1998**(1), 1–9 (1998)
13. Jacques, S.L.: Optical properties of biological tissues: a review. Physics in Medicine and Biology **58**(11), R37 (2013). URL http://stacks.iop.org/0031-9155/58/i=11/a=R37
14. Jolivot, R., Benezeth, Y., Marzani, F.: Skin parameter map retrieval from a dedicated multispectral imaging system applied to dermatology/cosmetology. Journal of Biomedical Imaging **2013**, 26 (2013)
15. Kaivosoja, J., Pesonena, L., Kleemolab, J., Pölönenc, I., Saloc, H., Honkavaarad, E., Saarie, H., Mäkynene, J., Rajalaa, A.: A case study of a precision fertilizer application task generation for wheat based on classified hyperspectral data from uav combined with farm history data. In: Proc. of SPIE Vol, vol. 8887, pp. 88870H–1 (2013)
16. Kim, R.H., Armstrong, A.W.: Nonmelanoma skin cancer. Dermatologic clinics **30**(1), 125–139 (2012)
17. Kubelka, P., Munk, F.: Reflection characteristics of paints. Zeitschrift fur Technische Physik **12**, 593–601 (1931)
18. Kuula, J., Pölönen, I., Puupponen, H.H., Selander, T., Reinikainen, T., Kalenius, T., Saari, H.: Using vis/nir and ir spectral cameras for detecting and separating crime scene details. In: SPIE Defense, Security, and Sensing, pp. 83590P–83590P. International Society for Optics and Photonics (2012)
19. Lomas, A., Leonardi-Bee, J., Bath-Hextall, F.: A systematic review of worldwide incidence of nonmelanoma skin cancer. British Journal of Dermatology **166**(5), 1069–1080 (2012)
20. Nascimento, J.M., Dias, J.M.: Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE transactions on Geoscience and Remote Sensing **43**(4), 898–910 (2005)
21. Neittaanmäki, N., Salmivuori, M., Pölönen, I., Jeskanen, L., Ranki, A., Saksela, O., Snellman, E., Grönroos, M.: Hyperspectral imaging in detecting dermal invasion in lentigo maligna melanoma. British Journal of Dermatology **177**(6), 1742–1744 (2017)

22. Neittaanmäki-Perttu, N., Grönroos, M., Jeskanen, L., Pölönen, I., Ranki, A., Saksela, O., Snellman, E.: Delineating margins of lentigo maligna using a hyperspectral imaging system. Acta dermato-venereologica **95**(5), 549–552 (2015)
23. Norvang, L., Milner, T., Nelson, J., Berns, M., Svaasand, L.: Skin pigmentation characterized by visible reflectance measurements. Lasers in Medical Science **12**(2), 99–112 (1997)
24. Palmadesso, J.B.P., Antoniades, J., Baumbeck, M., Rickard, L.: Use of filter vectors in hyperspectral data analysis. Infrared Spacebome Remote Sensing III, Proceedings of SPIE **2553**, 148–157 (1995)
25. Salmivuori, M., Neittaanmäki, N., Pölönen, I., Jeskanen, L., Snellman, E., Grönroos, M.: Hyperspectral imaging system in the delineation of ill-defined basal cell carcinomas: A pilot study. Journal of the European Academy of Dermatology and Venereology **33**(1), 71–78 (2019)
26. Stratigos, A., Garbe, C., Lebbe, C., Malvehy, J., Del Marmol, V., Pehamberger, H., Peris, K., Becker, J.C., Zalaudek, I., Saiag, P., et al.: Diagnosis and treatment of invasive squamous cell carcinoma of the skin: European consensus-based interdisciplinary guideline. European journal of cancer **51**(14), 1989–2007 (2015)
27. Trakatelli, M., Morton, C., Nagore, E., Ulrich, C., Del Marmol, V., Peris, K., Basset-Seguin, N.: Update of the european guidelines for basal cell carcinoma management. European Journal of Dermatology **24**(3), 312–329 (2014)
28. WHO: Skin cancers [www document]. http://www.who.int/uv/faq/skincancer/en/. Accessed on Dec 11,2018
29. Zheludev, V., Pölönen, I., Neittaanmäki-Perttu, N., Averbuch, A., Neittaanmäki, P., Grönroos, M., Saari, H.: Delineation of malignant skin tumors by hyperspectral imaging using diffusion maps dimensionality reduction. Biomedical Signal Processing and Control **16**, 48–60 (2015)