Multispectral Imaging Methods for the Diagnosis of Skin Cancer Lesions

Chenxue Liang

Supervised by Mr. Meritxell Vilaseca and Xana Delpueyo (CD6-UPC)

Presented on date 19th July 2016
Multispectral Imaging Methods for the Diagnosis of Skin Cancer Lesions

CHENXUE LIANG
Centre de Desenvolupament de Sensors, Instrumentació i Sistemes (CD6),
Universitat Politècnica de Catalunya (Terrassa - Barcelona, Spain).

E-mail: chobits4793@gmail.com

Abstract: Skin cancer is the most prevalent form of cancer, and melanoma is one of the most threat disease of it. But it can be cured if it is detected early enough. Multispectral imaging is a potential method to differentiate melanoma from nevi as it provides spectral images with information of absorbance and reflectance. With this aim, spectral images along the visible and near infrared range (from 415nm to 995nm) of 165 lesions including nevi, melanomas and basal cell carcinomas were processed in this master thesis. After obtaining all data in terms of reflectance and absorbance and other related parameters for each pixel of the segmented lesions, a statistical analysis was carried out to quantify their spatial distribution all over each lesion. Algorithms such as Support vector machine (SVM) and Discriminant Analysis (DA) were used as a means of classifying the lesions. The results show that DA linear classifier provides a better diagnosis than the SVM. BCCs are easier to discriminate from nevi than melanomas.

Keywords: melanoma, skin cancer, multispectral imaging system, reflectance, absorbance

1. Introduction

Skin cancer is the most prevalent form of cancer. More new cases of this cancer are diagnosed every year in comparison with breast, prostate, lung, and colon cancers [1]. Melanoma is a fatal disease that develops from the malignant transformation of melanocytes. This type of skin cancer is the one with worst diagnosis and prognosis and affects mostly adults, especially, for Caucasians in Europe and North America [2]. On the other hand, prevalence is equal for men and women. More than 90% of cases are caused by exposure to ultraviolet radiation from the Sun [3]. This exposure has increased partly due to a thinner ozone layer [4], and tanning beds are becoming another common source of UV radiation.

Skin cancer, and even melanoma, can be cured if they are detected early enough although it is a threat to human life. In fact, it is hard to distinguish melanomas from non-malignant nevi. Malignant lesions identified in the primary care setting are less than 50% [5] and as a consequence of the current diagnostic effectiveness, 27% of patients with melanoma die.

Different kinds of skin lesions have different reflectance and absorbance with light from different wavelengths as some published studies suggest. For instance, Kuzmina and colleagues [6] found that melanomas are linked to higher absorbance in the visible and near infrared, especially in the blue region (figure 1). Multispectral imaging (MSI) is a very useful non-contact, optical method for testing the reflectance and absorbance of samples with a precise spatial resolution [7]. For instance, using narrow-band filters at different wavelengths placed in front of a digital camera, a set of spectral images of the samples are taken [8]. Alternatively, light sources with narrow spectral emission, such as light-emitting diodes (LEDs) can be used to scan the sample along the different wavelengths. In any case, each pixel of the image contains information about the corresponding sample spot at a specific wavelength, and a complete reflectance spectrum for a single pixel can be obtained by processing together the set of multispectral images. Accordingly, MSI can be used in vivo with potentially promising applications to skin studies and pathology diagnosis. Specifically, spectral signatures linked to melanomas would be useful to distinguish them from normal nevi and some other skin cancers as basal cell carcinoma (BCC), which are less dangerous, as early as possible. This analysis is based on the fact that the penetration depth of the optical radiation in the skin tissues depends on wavelength. Also, the spectral
Multispectral imaging methods for the diagnosis of skin cancer lesions

reflectance/absorbance is dependent on tissue components (chromophores) such as hemoglobin, melanin, etc. and all this may vary among lesions of different etiology.

![Averaged optical density (OD) spectra for melanomas, other pigmented malformations, vascular malformations, and normal skin (Source: Kuzmina et al. 2011).](image)

However, the properties of melanoma in terms of reflectance and absorbance remain rather unknown and further research is still needed in this area to be able to differentiate them from other skin lesions. The main goal of this thesis is to find some useful parameters computed from reflectance/absorbance of lesions at different wavelengths, to improve the accuracy of skin cancer diagnosis.

This work is within the framework of the European Project DIAGNOPTICS “Diagnosis of skin cancer using optics” (ICT PSP seventh call for proposals 2013), the main goal of which is developing a multiphoton platform including hyperspectral and 3D techniques, blood flow analysis and confocal microscopy for in-vivo imaging of skin cancer lesions as a diagnosis service [9,10].

2. Experiment setup

2.1. Device

The device used in this master thesis is a handheld multispectral imaging system developed at the CD6 which includes a CCD camera attached to an objective lens and a ring light source with 8 types of LEDs with different spectral emissions within the visible and near infrared range (414, 447, 477, 524, 671, 735, 890, 995nm).

In order to obtain a uniform field of illumination on the sample and enough energy to acquire spectral images with low exposure times, 32 LEDs are included in the light source (4 of each type). Moreover, two polarizers allow changing the degree of polarization of light which allows obtaining information from different skin depths. Specifically, the first polarizer is located in front of the LEDs and the second in front of the objective lens. Light polarization can be changed into 3 different positions: 0° - parallel polarizers, 45° and 90° - crossed polarizers. The non-polarized light is used to obtain information from deeper layers under the skin surface as that polarized is mainly due to specular reflection from the surface of the skin. However, only the images acquired with crossed polarizers, and thus with information from deeper layers, have been used and processed in this study. Figure 2 (a) shows the whole device, while (b) and (c) show the calibrated reference inside the basic board, which will be used as the reflectance reference as it is explained below.
Multispectral imaging methods for the diagnosis of skin cancer lesions

With this device, a clinical study was previously conducted at the Hospital Clinic i Provincial de Barcelona (Spain) and multispectral images corresponding to 165 lesions (125 common nevi, 25 melanomas and 15 BCC) were obtained. In this master thesis these images have been processed.

2.2. Reflectance, absorbance, CIE L*a*b* and other related parameters

When the light impinges the skin, some of the light is reflected and the rest is absorbed or scattered. With the device described in the former section, multispectral images that contain information of the specular reflected light and back-scattered light at different wavelengths are obtained. However, it must be reminded that the specular component is mainly removed by using the crossed polarizers. From the multispectral images of the different samples acquired corresponding to the different spectral bands, the spectral reflectance can be computed as follows for each wavelength considered in the system (414, 447, 477, 524, 671, 735, 890, 995nm) assuming that the spectral bands are almost monochromatic:

\[
\rho = \frac{I - I_d}{I_c - I_d} \rho_r \quad (1)
\]

where \( \rho \) is the reflectance of the sample, \( I \) is the recorded intensity (digital level) in each pixel of the image of a sample, \( I_d \) and \( I_c \) are the recorded intensity (digital level) in each pixel of an image of calibrated reference and a dark current image recorded without any sample, respectively and \( \rho_r \) is the reflectance of the calibrated reference. A gray patch (Neutral 6.5 from the X Rite ColorChecker® Classic CCCR) with known spectral reflectance in the visible and near infrared was used as a reference.

The absorbance at each wavelength can also be easily computed from the reflectance as follows:

\[
A = \log_{10} (\rho) \quad (2)
\]

Once this is done, the system computes both the reflectance and the absorbance every 10nm in the analyzed range by interpolating those initially available.

A Matlab software to perform such calculations is also available at the CD6. It allows selecting a squared area on an image of a lesion or alternatively doing its segmentation, obtaining the reflectance for each pixel and also the mean (± standard deviation values). In figure 3 it can be seen that the reflectance value of skin is always higher than that of the lesion, which looks darker and therefore absorbs more light.
Multispectral imaging methods for the diagnosis of skin cancer lesions

Besides reflectance and absorbance, other parameters can be computed which can be useful to enhance spectral differences among different types of skin lesions. They are basically based on absorption peaks in the visible and the near infrared and of skin chromophores which can be different in healthy and malignant tissue and might be useful to get a mapping from them and thus to improve diagnosis. Some of them are internal hemoglobin, bilirubin, melanin and so on. The following have been found in the literature [11,12]:

\[ H = \frac{I_{671}^2}{I_{524}^2} \]  
\[ B = \frac{I_{447}}{I_{671}} \]  
\[ M1 = \frac{I_{995}}{I_{524}} \]  
\[ M2 = \frac{(I_{671} \cdot I_{995})}{I_{524}^2} \]  
\[ P = \frac{I_{524}}{(I_{671} \cdot I_{995})} \]  
\[ \log(P) = \log\left(\frac{I_{524}}{I_{671} \cdot I_{995}}\right) \]  
\[ E = \frac{I_{671}}{I_{524}} \]  
\[ O = \frac{I_{671}}{I_{995}} \]  
\[ Test = \frac{I_{524}}{I_{447} \cdot I_{890}} \]

\( I_{\lambda} \) is the intensity recorded at a specific wavelength. The equations can be useful to describe some chromophores of the tissue: (3) Hemoglobin, (4) Bilirubin, (5) Melanin1, (6) Melanin2, (7) Melanoma index, (8) Melanoma index in logarithmic units, (9) Erythema index, (10) Oximetry (11) Experiment test.

Furthermore, the software shown in figure 3 also processes some colorimetric values (CIEL*a*b*) pixel by pixel, which can be though as a three-dimensional real number space that contains an infinite possible representations of colors [13]. From figure 4, the lightness, \( L^* \), represents the darkest black at \( L^* = 0 \), and the brightest white at \( L^* = 100 \). The color channels, \( a^* \) and \( b^* \), will represent true neutral gray values at \( a^* = 0 \) and \( b^* = 0 \). The red/green opponent colors are represented along the \( a^* \) axis, with green at negative \( a^* \) values and red at positive \( a^* \) values. The yellow/blue opponent colors are represented along the \( b^* \) axis, with blue at negative \( b^* \) values and yellow at positive \( b^* \) values. On the other hand, the colorimetric values Chrome (C*) and Hue (h) were also computed (as in (12), (13)).

\[ C_{ab} = \sqrt{a^2 + b^2} \]  
\[ h_{ab}^o = \arctan\left(\frac{b^*}{a^*}\right) \]
Multispectral imaging methods for the diagnosis of skin cancer lesions

Chroma refers to the perceived saturation of the color, i.e., if the color appears vivid or contains a lot of white, and hue corresponds to the color angle in the CIEL*a*b* color space from 0° to 360°, i.e., red (0°), yellow (90°), green (180°), blue (270°) to red again (360°).

Taking into account all the former information, images containing reflectance and absorbance values as well as the other parameters described (pixel by pixel) have been computed.

2.3 Boxplots with Energy, Entropy and Third central moment

After obtaining all the data of reflectance and absorbance and related parameters for each pixel of the segmented lesions, their mean and standard deviation (SD) have been computed.

Additionally, a statistical analysis has been carried out to quantify their values over the pixels of each lesion. This analysis might provide spatial information about the distribution of colors or spectral features all over the sample. To do so, the histogram of the pixel values within the segmented lesion in terms of all computed parameters is firstly plotted. Then, a boxplot like that in figure 5 is also plot, which is a standardized way of displaying the distribution of data based on the five number summary: minimum value, first quartile (25%), median, third quartile (75%), and maximum value. The line extending vertically from the boxes (whiskers) indicates variability outside the upper and lower quartiles, and outliers may be plotted as individual points.

Hence, the spacing between the different parts of the box indicate the degree of dispersion (spread) and skewness in the data, and show outliers. Once this is available, some statistical descriptors have been implemented in order to account for some features of the histogram shape, such as entropy (Ep), which gives information about randomness of values, energy (En) for the uniformity, and the third central moment (µ3), which accounts for the skewness of the histogram (as in (14), (15) and (16)), in the interest of having a visual conception and a first approach to the differences among benign and malignant lesions.

\[
E_p = \sum_{i=0}^{n-1} P_i \log_2(P_i) \\
E_n = \sum_{i=0}^{n-1} P_i^2 \\
\mu_3 = \sum_{i=0}^{n-1} (i - m)^3 P_i
\]

where \( P_i \) is the value of the element \( i \) of the histogram, \( n \) is the number of levels that the histogram is divided into and \( m = \sum_{i=0}^{n-1} i P_i \) is the mean (average) value.

Figure 5. Boxplot
Multispectral imaging methods for the diagnosis of skin cancer lesions

2.4. Support Vector Machine (SVM) and Discriminant Analysis (DA)

From the former section, we obtained 80 descriptors about the data of reflectance and absorbance including the mean, SD, En, Ep and $\mu_3$ at 8 wavelengths), 45 descriptors of the 9 other related parameters based on chromophore absorption (Hemoglobin etc.) and 5 based on color metrics ($L^*$, $a^*$, $b^*$ etc.). As the final goal of this thesis is improving skin cancer diagnosis, the use of classification algorithms such as Support vector machine (SVM) and Discriminant Analysis (DA) is proposed. The parameters described above are used as the input data for these algorithms.

Classifying data is a common task in machine learning; Specifically, SVM is useful to distinguish between only 2 classes; Suppose that some given data points belong to one of two classes, and the goal is to decide which class a new data point will be in. In the case of SVM, a data point is viewed as a p dimensional vector, and a p-1 dimensional hyperplane is used to separate such points into the 2 classes.

On the other hand, DA is a statistical analysis method to select a criterion to determine how to place a new sample between two or more naturally occurring classes. It assumes that different classes generate data based on a different Gaussian distribution. The advantage of this classifier in comparison to SVM is that more than 2 classes can be considered simultaneously. Like the SVM, each the data point of lesion is viewed as p dimensional vector, and we use the Gaussian distribution to separate such points with a p-1 dimensional hyperplanes. All the data of nevi, melanoma and BCC will be distributed as a Gaussian function in a same plane and each kind of lesion has its own Gaussian distribution function (figure 6).

In order to compute the accuracy of the classification algorithms (SVM and DA) a confusion table or matrix is used [14]. This matrix is a specific table layout that allows visualization of the performance of an algorithm for the classification of samples into several classes. Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class and vice-versa [15] (figure 7).

Assuming p represents Class 1 (for instance, melanoma – malignant lesion) and n represents Class 2 (for instance, nevus – benign lesion), the confusion table shows: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). Then, the accuracy (ACC) and the error rate in each classification (SVM and DA) is obtained by the following calculation: $\text{ACC} = (TP + TN) / (#p + #n)$, $\text{Error} = 1 - \text{ACC}$.

In this study, the former algorithms have been use to classify 3 classes of samples: Nevi, melanoma and BCC. In the case of SVM several comparisons including only 2 classes have been done: Nevi-Melanoma, Nevi-BCC, Melanoma-BCC. In the case of DA the former comparisons have been included in the algorithm as well as all three classes at once.

Both in SVM and DA, the confusion table and error have been computed when:

1. All samples are included for training and test the classifier.
2. All samples are included for training except one which is used for testing the classifier. This process is repeated by all samples and a mean error is calculated.
3. Half of the samples are randomly selected for training and the other half are used to test the classifier. This is repeated 20 times and a mean error is calculated.

3. Results

Figure 8 shows the spectral images within the visible and near infrared range taken with the multispectral system. From all the lesions analyzed, images like the former ones were available and they were segmented.
Multispectral imaging methods for the diagnosis of skin cancer lesions

As described in the former section, the mean (± SD) in terms of reflectance and absorbance at the 8 different wavelengths were computed as well as the other parameters (hemoglobin, melanin, etc.) and CIEL*a*b* coordinates. Then the histograms were plotted and related descriptors (energy, entropy and third central moment) for each of the parameters were also calculated, thus obtaining the distribution of nevi, melanomas and BCC (which represent the classes).

Some of the best boxplots obtained with the descriptors (energy, entropy and third central moment values) calculated for all the lesions processed are shown in figure 9. The results plotted correspond to the descriptors which allow a better visual discrimination among types of lesions. It can be seen that in average, the En at 975nm and 995nm of nevi are higher than the melanoma and BCC, the Ep of melanoma at 715nm and μ3 at 735nm are larger than nevi and BCC, and the En of nevi at 575nm is lower than the melanoma and BCC. This is probably due to the fact that nevi are more uniform than melanoma and BCC in terms of color and spectral features.

Even though it can be seen that parameters from the three types of lesions slightly overlap, a clear difference in the mean values start to become evident when comparing melanomas with the other two types of lesions even the limited number of samples analyzed.

<table>
<thead>
<tr>
<th>Energy Reflectance 575nm</th>
<th>Energy Reflectance 995nm</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Boxplot" /></td>
<td><img src="image2.png" alt="Boxplot" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entropy Reflectance 715nm</th>
<th>μ3 Absorbance 735nm</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.png" alt="Boxplot" /></td>
<td><img src="image4.png" alt="Boxplot" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Energy Reflectance 975nm</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5.png" alt="Boxplot" /></td>
</tr>
</tbody>
</table>

In the tables below, the best results obtained with SVM and DA are shown. Table 1 shows the classification (confusion matrix and error) when considering conditions 1, 2 and 3 described in the former section when the
Multispectral imaging methods for the diagnosis of skin cancer lesions

5 descriptors for the 8 Reflectance values are used as inputs to discriminate melanomas from nevi. Table 2 shows the results but when considering 5 descriptors for the 8 Reflectance and Absorbance values, the related chromophore-based parameters (hemoglobin, melanin, etc.) are used as inputs to discriminate BCC from Nevi. Table 3 shows the classification of 3 classes (Nevi, Melanomas and BCCs) when considering the 8 Reflectance and Absorbance values as inputs (a), and the 5 descriptors for the 8 Reflectance and CIEL*a*b* parameters (b). AV, Min and Max mean the average, minimum and maximum value of error.

Table 1. Confusion matrices and corresponding errors using SVM and DA when 2 classes are considered: Nevi and Melanoma, in which 5 descriptors for the 8 Reflectance values are used as inputs. AllLesions refer to condition number 1, Leaveout one to number 2 and Leaveout half to number 3 (see subsection 2.4).

<table>
<thead>
<tr>
<th>SVM</th>
<th>Discriminant Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AllLesions</td>
<td>Error</td>
</tr>
<tr>
<td>124 1</td>
<td>0.1333</td>
</tr>
<tr>
<td>19 6</td>
<td></td>
</tr>
<tr>
<td>124 1</td>
<td>0.1467</td>
</tr>
<tr>
<td>21 4</td>
<td></td>
</tr>
<tr>
<td>123 2</td>
<td>AV: 0.1593</td>
</tr>
<tr>
<td>20 5</td>
<td>Min: 0.1333 Max:0.1867</td>
</tr>
</tbody>
</table>

Table 2. Confusion matrices and corresponding errors using SVM and DA when 2 classes are considered: Nevi and BCC, in which 5 descriptors for the 8 Reflectance and Absorbance values and the related chromophore-based parameters (hemoglobin, melanin, etc.) are used as inputs. AllLesions refer to condition number 1, Leaveout one to number 2 and Leaveout half to number 3 (see subsection 2.4).

<table>
<thead>
<tr>
<th>SVM</th>
<th>Discriminant Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AllLesions</td>
<td>Error</td>
</tr>
<tr>
<td>108 17</td>
<td>0.2</td>
</tr>
<tr>
<td>11 4</td>
<td></td>
</tr>
<tr>
<td>101 24</td>
<td>0.2357</td>
</tr>
<tr>
<td>9 6</td>
<td></td>
</tr>
<tr>
<td>73 52</td>
<td>AV: 0.3843 Min: 0.1847 Max:0.6</td>
</tr>
</tbody>
</table>

Table 3. Confusion matrices and corresponding errors using DA when 3 classes are considered (nevi, melanoma and BCC). AllLesions refer to condition number 1, Leaveout one to number 2 and Leaveout half to number 3 (see subsection 2.4). (a) corresponds to the results obtained when the 8 Reflectance and Absorbance values are considered as inputs. And (b) is obtained when the 8 Reflectance values and CIEL*a*b* parameters are used as inputs.

<table>
<thead>
<tr>
<th>(a) Discriminant Analysis</th>
<th>(b) Discriminant Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AllLesions</td>
<td>Error</td>
</tr>
<tr>
<td>124 1</td>
<td>0.0303</td>
</tr>
<tr>
<td>4 21 0</td>
<td>3 22</td>
</tr>
<tr>
<td>0 0 15</td>
<td></td>
</tr>
<tr>
<td>102 16</td>
<td>0.3091</td>
</tr>
<tr>
<td>16 6 3</td>
<td>18 4</td>
</tr>
<tr>
<td>7 2 6</td>
<td>5 5</td>
</tr>
<tr>
<td>44 42 39</td>
<td>AV: 0.59</td>
</tr>
<tr>
<td>8 9 8</td>
<td>Min: 0.4364</td>
</tr>
<tr>
<td>4 8 3</td>
<td>Max:0.7818</td>
</tr>
</tbody>
</table>
Multispectral imaging methods for the diagnosis of skin cancer lesions

From the results it can be seen that when all samples are used as training set, DA performs better than SVM; and this is true also in the other 2 additional conditions considered, i. e., when all samples are included for training except one which is used for testing, and when half of the samples are randomly selected for training and the other half are used for testing (see in 2.4).

In addition, for these two additional conditions, the errors of SVM are slightly lower than those of DA; however, it is worth noting that DA can distinguish more melanomas than the SVM (Table 1, conditions 2 and 3). Furthermore, from table 2 it can be seen that in general, the classification of nevi and BCC is better than that of the nevi and melanoma. This is because BCC has more differentiated spectral features from nevi than melanoma.

Table 3 shows that when the 3 samples are considered together, the error increases, especially when half of the samples are used as training and the rest for testing.

4. Conclusion

In this study we used multispectral images previously taken from a multispectral device to get the spectral information of skin lesions. Images from 165 lesions were considered including 125 nevi, 25 melanomas and 15 BCCs. Then, we used a Matlab program to segment each of the lesions and calculated reflectance and absorbance values and related parameters based on them. Boxplots were graphed and then, a statistical analysis was carried out to quantify their values over the pixels of each lesion. We calculated the entropy (Ep), energy (En), and the third central moment (µ3). It was shown that several specific wavelengths are more appropriate to discriminate among skin lesions of different types, i.e., nevi, melanoma and BCC. This can probably be attributed to the fact that nevi are more uniform than melanoma and BCC in terms of color and spectral features.

The SVM algorithms allowed classifying lesions with an error that ranged from 0.03% to 78%, meanwhile, the DA was link to an error that ranged from 0% to 52%, depending on the algorithm, parameters (Reflectance/Absorbance/Other) and classes (Type of lesions) considered. According to results of former section, we concluded that DA linear classifier provided a better diagnosis than SVM. Although the table 1 shows the errors of SVM lower than the DA, the confusion table of DA can distinguish more melanomas than the SVM. In addition, BCCs were easier to classify from nevi than melanomas. Furthermore, the results when comparing 2 classes are better than when comparing 3 of them, which means that the classification of 3 classes is more difficult. Only 25 melanomas and 15 BCCs were included in the data so that it is though that the detection error rate could be much reduced if more lesions were included. This will be done in future work.

Acknowledgments

This research was supported by the European Commission of Research & Innovation under the grant DIAGNOPTICS “Diagnosis of skin cancer using optics2 (ICT PSP seventh call for proposals 2013, 2014-2016). I am indebted to my advisor, Prof. Meritxell Vilaseca for giving me the opportunity of doing this thesis and Xana Delpueyo for her help and encouragement throughout the course of this work.

References

Multispectral imaging methods for the diagnosis of skin cancer lesions


[10] Delpueyo X; Vilaseca M; Royo S; Ares M; Sanabria F; Herrera J; Burgos FJ; Pujol J; Puig S; Pellacani G; Vázquez J; Solomita G; Bosch T. "Handheld hyperspectral imaging system for the detection of skin cancer". 2015. In the Midterm Meeting of the AIC (AIC 2015) and 16th International Symposium on Multispectral Color Science (MCS 2015), *Proceedings*, Japan. 385-390.


