ESTIMATION OF AN OPTIMAL SPECTRAL BAND COMBINATION TO EVALUATE SKIN DISEASE TREATMENT EFFICACY USING MULTI-SPECTRAL IMAGES

Sylvain Prigent¹, Didier Zugaj², Xavier Descombes¹, Philippe Martel², Josiane Zerubia¹

¹EPI Ariana, INRIA/I3S, 2004 route des Lucioles, BP93, 06902 Sophia Antipolis, Cedex, France
Sylvain.Prigent@inria.fr, Xavier.Descombes@inria.fr, Josiane.Zerubia@inria.fr
²Galderma R&D, 2400 Route des Colles, BP87, 06902 Sophia-Antipolis, Cedex, France
Didier.Zugaj@galderma.com, Philippe.Martel@galderma.com

ABSTRACT

Clinical evaluation of skin treatments consists of two steps. First, the degree of the disease is measured clinically on a group of patients by dermatologists. Then, a statistical test is used on obtained set of measures to determine the treatment efficacy. In this paper, a method is proposed to automatically measure the severity of skin hyperpigmentation. After a classification step, an objective function is designed in order to obtain an optimal linear combination of bands defining the severity criterion. Then a hypothesis test is deployed on this combination to quantify treatment efficacy.

Index Terms— hypothesis tests, t-test, multispectral, skin, hyperpigmentation

1. INTRODUCTION

One of the steps to evaluate the efficacy of a therapeutic solution is to perform measurements on a series of patients who received the studied treatment. In parallel another treatment is tested on another group of people or on the same group of patients on another skin area. This second treatment is the reference one for the studied pathology or a placebo. In the following we will use the word ‘vehicle’ to refer to this treatment. For facial hyper-pigmentation, for each studied treatment, a group of patients by dermatologists. Then, a statistical test is deployed to measure if the severity of skin hyperpigmentation evolves significantly. The used test is a t-test ([6] which quantifies the deviation of the mean value between two distributions:

\[ Z^{t,t_0} = \frac{X(t) - X(t_0)}{\sqrt{\frac{\sigma^2(t)}{N_c} + \frac{\sigma^2(t_0)}{N_c}}} \]

where \( X \) and \( \sigma \) are the mean value and the standard deviation of the two distributions respectively. In our application, a distribution is set of severity measurements on \( N_c \) patients. The t-test is performed between a measure at time \( t \) and at the baseline at the time \( t_0 \). The null hypothesis is: “The mean
value of the distribution does not evolve from $t_0$ to $t^*$. If the probability of false alarm (p-value) of the test is below 0.05 the null hypothesis is rejected. And, the lower the p-value, the higher statistically is the deviation between the mean values of the two distributions, and the treatment is proved to be efficient.

The paper is organised as follows. The second section defines the normalization of the data to measure the disease severity. The third section describes the objective function proposed to find an optimum severity criterion per treatment. Then the fourth section shows the obtained results and a comparison with the benchmark method.

2. DATA AND NORMALISATION

For each treatment, the data set of a study is composed of $N_e$ patients. Each patient has two treatments, the active product and the vehicle, in two different areas. Thus there are two multispectral images per patient. After classification of healthy and pathological areas, an average spectrum is computed for each region. Then, four spectra are obtained per patient and per time measurement: the healthy and pathological spectra for the active product and for the vehicle. It comes that two normalizations are needed to compute a severity value from a criterion $M$. The first one is computed to normalize the criterion measurement between healthy and pathological area, and the second to normalize the measurement of the active treatment by the vehicle. For the healthy/pathological normalization there are two possibilities:

$$d_t^e = \mu_{M_h} - \mu_{M_p}$$  \hspace{1cm} (3)

and

$$d_t^e = \frac{\mu_{M_h} - \mu_{M_p}}{\mu_{M_h}},$$  \hspace{1cm} (4)

where $d_t^e$ denotes the severity of the disease for the patient $e$ at the measurement time $t$, and $\mu_{M_h}$ and $\mu_{M_p}$ are respectively the average measurement of the darkness due to the criterion $M$ on the healthy and the pathological area respectively. The equation (4) seems to be the best choice as it computes a deviation percentage of the pathological area with respect to the healthy one. Nevertheless, the quotient introduces additive noise. And as the two quantities $\mu_{M_h}$ and $\mu_{M_p}$ are homogeneous the normalization defined in equation (3) is preferred.

A second normalization is needed to compare the active treatment and the vehicle. With the two images taken from the same patient, there is again the two possible normalizations:

$$D_t^e = d_t^e, A - d_t^e, V$$  \hspace{1cm} (5)

and

$$D_t^e = \frac{d_t^e, A - d_t^e, V}{d_t^e, V},$$  \hspace{1cm} (6)

where $d_t^e, A$ and $d_t^e, V$ are the severities measured at the time $t$ on the patient $e$ on the image concerned by the active treatment $A$, and the vehicle treatment $V$ respectively. For the same reason as above, the chosen normalization is the one of equation (5). Then the final measurement of the severity according to the studied treatment $A$ compared to the reference treatment $V$ is:

$$D_t^e = (\mu_{M_h,A} - \mu_{M_p,A}) - (\mu_{M_h,V} - \mu_{M_p,V})$$  \hspace{1cm} (7)

It can be noticed that this normalization is linear.

3. THE OBJECTIVE FUNCTION

To determine the severity $D_t^e$ of the patient $e$ at the time $t$, a criterion $M$ is needed. In this section, we propose to define an objective function to automatically estimate a criterion $M$ which is optimal to describe the treatment effects. If a treatment has an effect, the deviation between the healthy and the pathological areas decreases from the measurement times $t$ to $t + dt$. So, the best criterion will be the one which describes this evolution with the finest scale. This is why, we design an objective function that maximises both the deviation between healthy and pathological areas and the deviation of the severity between time measurements:

$$\Delta_{t\rightarrow t_0} \left[ \frac{1}{N_e} \sum_{e=1}^{N_e} [D_t^e, M] \right],$$  \hspace{1cm} (8)

where $\Delta$ is the differential between the initial time $t_0$ and the time $t$:

$$\Delta_{t\rightarrow t_0} (X_t) = X_{t_0} - X_t.$$  \hspace{1cm} (9)

Since the objective is to maximize the deviation mentioned above for any measurement time $t$ with respect to the criterion $M$, the optimisation problem can be summarized by:

$$\hat{\lambda} = \arg \max_{\lambda} f(\lambda),$$  \hspace{1cm} (10)

where $f$ is the objective function defined by:

$$f(\lambda) = \sum_{t=t_0 + dt}^{N_t} \left\{ \Delta_{t\rightarrow t_0} \left[ \frac{1}{N_e} \sum_{e=1}^{N_e} [D_t^e, \lambda] \right] \right\}$$  \hspace{1cm} (11)

where $\lambda$ is the vector of coefficients defining the criterion $M$ and $N_t$ the number of time measures. To optimize the expression given by equation (10), the geometric simplex method is used [7].

As the final value which quantifies the efficacy of a treatment comes from a $t$-test one can propose to directly minimize the p-value of the $t$-test. Nevertheless, the obtained spectrum will be correlated neither with the disease, non with the treatment effect. That’s why we prefer not to do so.
4. EXPERIMENTAL RESULTS

4.1. The data set

The used data are 384 multispectral images taken from 48 patients during 3 months with one measure at the first visit and one every month. Multi-spectral images contain 18 bands in the visible and near infra-red spectra. The group of 48 patients is partitioned in 3 sub-groups of 16 patients. Each sub-group corresponds to a specific tested treatment.

4.2. The treatments

The first treatment is the benchmark for skin pigmentation. We will note it $S_t$. The two other treatments are from the same product with two different doses. It is the tested product. The two treatments will be called $A_d2$ and $A_d3$ where $d$ stand for dose and $d1 < d2$. $A$ is written for Active. For the vehicle, we use the active product with a low dose. We will note it $A_d1$ with $d1 < d2$. Thus on the 3 groups we measure the treatments $S_t \; vs \; A_d1$, $A_d2 \; vs \; A_d1$ and $A_d3 \; vs \; A_d1$.

4.3. The baseline

To analyse our method, we use the $L^*$ measure from the CIEL*a*b. To obtain $L^*$, a spectral integration algorithm projects the multi-spectral image to the XYZ colours space. Then the transformation from XYZ space to CIEL*a*b space allows us to extract the $L^*$ component.

4.4. The results

To avoid some bias due to the classification algorithm, pathological and healthy areas are manually selected on each image. Those areas are manually registered to be in correspondence for each patient along the time sequence. Fig. 2 shows the p-values obtained by the criterion computed with the optimisation method. P-values below 0.05 are highlighted with bold. The tested product with a $d1$ dose has an efficacy equivalent to the standard treatment $S_t$. The null hypothesis of the t-test is accepted with a high p-value for all the time sequences. When the dose of the product $A$ increases, it effect should increase and the higher the dose, the longer the effect. The p-values of the t-test with the doses $A_d2$ and $A_d3$ describe well this behaviour. The criterion is adapted to quantify hyper-pigmentation. Fig. 3 summarizes the p-value of the same test but with the Luminance ($L^*$) as a criterion. This test shows despite that $S_t$ and $A_d1$ are equivalent, for higher doses of $A$, the p-values do not reflect as well as previously the treatment effect. Indeed, for the first measurement, after one month, the effect is not visible; the p-values are above 0.05. Then, along the time sequence, even if the p-values for the treatment $A_d3$ follow the expected variations, the p-values variations for $A_d2$ are less coherent since they decrease under the 0.05 threshold and then goes back above this threshold. Thus, the conclusions are that the proposed criterion obtained by optimisation is more sensitive to small variations than the standard criterion $L^*$. To illustrate this sensitivity, we show in Fig. 1 an example of the evolution of one patient that dermatologists estimate cured at the end of the treatment. As one can see, the measurement follows precisely the disease evolution.

We propose now an interpretation of the criteria. The Luminance is obtained from the component $Y$, by spectral integration on the multi-spectral images. As a result, $L^*$ in the proposed process is equivalent to use the bands combination $Y$ shown in Fig. 5. Fig. 4 represents the $\lambda$ coefficients depending on the wavelength obtained by the proposed method. All the obtained criteria have similar shapes but with small variations due to the optimality to each treatment. If we compare the obtained criterion curves with the absorbency curves of melanin and haemoglobin (see Fig. 6) it can be noticed that there is a link between the two. In fact, coefficients of linear combinations are high, in absolute value, when the deviation between haemoglobin and melanin absorbency is high. This observation explains why the proposed method is more efficient to describe the effect of a treatment. When $L^*$ measures only the middle of the spectrum, the objective function allows to select adapted spectral bands.

![Fig. 1](image1.png)

$S_t \; vs \; A_d1$ with $t_{027} = 10^{-1}$, $t_{282} = 10^{-1}$, $t_{379} = 10^{-1}$.

$A_d2 \; vs \; A_d1$ with $t_{1158} = 10^{-2}$, $t_{4379} = 10^{-8}$, $t_{3573} = 10^{-3}$.

$A_d3 \; vs \; A_d1$ with $t_{2240} = 10^{-2}$, $t_{5282} = 10^{-8}$, $t_{5458} = 10^{-4}$.

**Fig. 2.** P-values obtained by minimizing the objective function on a spectrum from 400 to 700 nm.
<table>
<thead>
<tr>
<th></th>
<th>$t_1 - t_0$</th>
<th>$t_2 - t_0$</th>
<th>$t_3 - t_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_t$ vs $A_{d1}$</td>
<td>$9.209 \times 10^{-1}$</td>
<td>$5.977 \times 10^{-1}$</td>
<td>$9.071 \times 10^{-1}$</td>
</tr>
<tr>
<td>$A_{d2}$ vs $A_{d1}$</td>
<td>$2.577 \times 10^{-2}$</td>
<td>$1.478 \times 10^{-2}$</td>
<td>$8.625 \times 10^{-2}$</td>
</tr>
<tr>
<td>$A_{d3}$ vs $A_{d1}$</td>
<td>$1.072 \times 10^{-1}$</td>
<td>$4.833 \times 10^{-3}$</td>
<td>$2.996 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

**Fig. 3.** P-values obtained with the Luminance ($L^*$) as the darkness measure.

**Fig. 4.** Coefficients of bands computed by the minimisation of the objective function for every couple of tested treatments.

**Fig. 5.** Coefficient of the linear combination of spectral band to obtain the XYZ decomposition depending on the wavelength.

**Fig. 6.** Absorbency curves of melanin and haemoglobin depending on the wavelength.

5. **CONCLUSION**

In this paper, we have proposed a method to automatically evaluate the efficacy of treatments with multi-spectral images. The optimization of an objective function allows to find an optimal spectral metric for each tested treatment. We show that such a metric gives a better analysis of the treatments than the standard method. In future work, the spatial information will be studied. In fact, only the spectral information of the average value of a lesion has been exploited. The spatial distribution can give information on the disease and its evolution.

6. **ACKNOWLEDGEMENT**

The authors would like to thank Dr. Laurent Petit and Anne-Sophie Dugaret for interesting discussions, and Galderma R&D early development for partial funding and for providing the data.

7. **REFERENCES**


