

# A Colour Image Processing Method for Melanoma Detection

O. Colot<sup>1</sup>, R. Devino<sup>2</sup>, A. Sombo<sup>2</sup>, D. de Bruca<sup>2</sup>

<sup>1</sup>PSI-LCIA, INSA de Rouen, Place E. Blondel, 76131 Mont-Saint-Aignan Cédex, France.

<sup>2</sup>PSI-La3I, Université Rouen, Place E. Blondel, 76821 Mont-Saint-Aignan Cédex, France.  
Olivier.Colot@insa-rouen.fr

## Abstract.

In this paper, we propose a method dedicated to classification between benign and malignant lesions in Dermatology in the aim to help the clinicians for melanoma diagnosis<sup>1</sup>.

The proposed methodology reduces the very numerous informations contained in the digitized images to a finite set of parameters giving a description of the colour and the shape of the lesions.

The whole process is shared in three steps: preprocessing, segmentation and classification of the lesions.

The proposed method was applied on a data base of 38 lesions (20 benign lesions and 18 malignant lesions) in the aim to assess the feasibility of the proposed method. The good classification rate obtained with the method is discussed and later tests to engage are underlined.

## 1. Introduction

Melanoma is an increasing form of cancer in the world. It has increased twice times for 15 years in Canada and it is 3% in the USA at this time.

The rate of clinicians visual investigations gets at 65% of good detection at the very best. In particular, it is very difficult to distinguish some atypical lesions - which are benign - from melanoma because they have the same properties according to the well known ABCDE rules used by dermatologists [1].

There is a visual inspection problem for the atypical class. The unnecessary excision is often practise for these lesions.

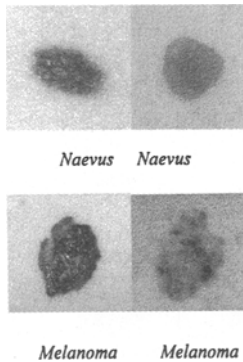
The variability of colours and shapes can lead to different interpretation by different dermatologists. So, there is a variability and a no reproductibility of the diagnosis. Different lesions are given in Figure 1.

However, melanoma is well suited for image processing because it is on the skin. Some researches [2-9] have shown the advantages to use image processing in dermatology.

---

<sup>1</sup> This work is supported by the french *Ligue Nationale contre le Cancer*.

Furthermore, the image processing by computer ensures the reproducibility of the analysis. However, the essential difficulty is to design robust and relevant parameters to ensure the separation between melanoma and benign lesions, in particular the atypical lesions (benign) which can be clinically mistaken for melanoma.



**Fig. 1. Some lesions**

At first, the lesion border is the first feature to identify. It is the first step of the processing to engage in order to extract informations about the lesion. So, the border extraction or identification is a critical step in computerized vision analysis in skin cancer as pointed out in [2]. Then, the segmentation step which takes place in any classification processing has to be really accurated.

As discussed in an early review [8], *computer image analysis in skin cancer is in its infancy*. The different developed approaches give results which are more or less accurate, efficient or reliable. It appears that in general each method needs a standardized image acquisition in order to be efficient. However, that is theoretical. Even you take care with the image acquisition conditions, there always exists deviations from standard conditions you define and you need to take them into account in the late image processing. We precise the preprocessing rules of our approach in section 2 where the deviations in image acquisition are taking into account and corrected before the image segmentation step.

The segmentation of coloured images does not involve as numerous studies as grey level image segmentation. In [4] an automatic colour segmentation algorithm is proposed for identification of skin tumor based on the use of six colour spaces. They apply a principal component transform in order to perform the border identification. We develop a multi-model approach too, but with different colour models leading to the definition of a 14-dimensional model of a colour image. This is proposed in section 3.

The whole process is then closed with the classification step developed in section 4.

The dermatologists use slides for lesion images storage and visual inspection for comparison. So, we based our processing on the use of such an information material. We digitalize the slides with a 35mm film scanner Nikon LS-1000.

## 2. Preprocessing

The preprocessing step is necessary because the images are noisy, the lighting conditions and the acquisition conditions are quite different from one slide to another. Furthermore, the images contains bright pixels due to local lighting conditions and local photometric properties of the lesions. So, we have developed a preprocessing technique which allows a photometric fitting in the aim to improve the images before the segmentation and the classification steps. The colorimetric fitting model or photometric corrective model we propose allows to have the safe skin around the lesion as close to a reference as possible.

In the aim to preserve the dynamics of the R, G and B values, we use a linear transformation on the safe skin area and on the lesion area.

The fitting model is written as it follows:

$$\begin{cases} R' = \frac{R - \overline{R}_S}{\sigma_{RS}} \times \sigma_{Rref} + \overline{R}_{Sref} \\ G' = \frac{G - \overline{G}_S}{\sigma_{GS}} \times \sigma_{Gref} + \overline{G}_{Sref} \\ B' = \frac{B - \overline{B}_S}{\sigma_{BS}} \times \sigma_{Bref} + \overline{B}_{Sref} \end{cases} \quad (1)$$

where  $R'$  (respectively  $G'$  and  $B'$ ) is the transformation of a  $R$  value of one pixel,  $\overline{R}_S$  (respectively  $\overline{G}_S$  and  $\overline{B}_S$ ) is the mean of  $R$  in the safe skin to be fitted,  $\overline{R}_{Sref}$  (respectively  $\overline{G}_{Sref}$  and  $\overline{B}_{Sref}$ ) is the mean of  $R$  in the reference safe skin  $\sigma_{RS}$  (respectively  $\sigma_{GS}$  and  $\sigma_{BS}$ ), is the standard deviation of  $R$  in the safe skin to be fitted and  $\sigma_{Rref}$  (respectively  $\sigma_{Gref}$  and  $\sigma_{Bref}$ ), is the standard deviation in the reference safe skin.

Figure 2 shows the results of the colorimetric corrective model applied to a same lesion taken with three different acquisition conditions.

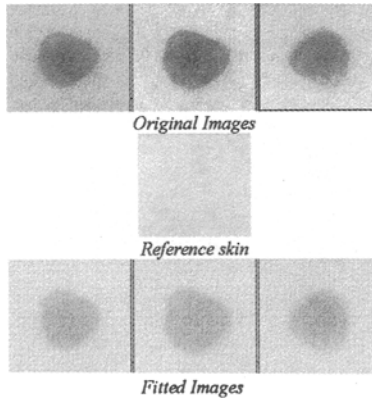


Fig. 2. Photometric correction

### 3. Segmentation

So as to obtain geometric and colorimetric informations on a lesion, it is necessary to run a segmentation process which will allow to extract the pixels belonging to the lesion from the image. The segmentation process we have developed is based on the cooperation between a segmentation in regions and a segmentation based on edge detection [10]. So, we consider there exists two kinds of pixels in the image:

1. pixels which belong to the safe skin
2. pixels which belong to the lesion.

So, we consider the image is made of two populations. The problem can then be viewed as a problem of two-hypotheses test:

1. Hypothesis  $H_S$ : the pixel  $(i,j)$  is a pixel in the safe skin
2. Hypothesis  $H_L$ : the pixel  $(i,j)$  is a pixel in the lesion

where  $i$  and  $j$  are the coordinates such as  $i \in \{1, \dots, N\}$  and  $j \in \{1, \dots, M\}$  and  $N \times M$  is the image dimension. Under these considerations, we compute a threshold obtained by means of the use of the *Entropy Maximization Principle* (EMP) [11] that is the coarse segmentation. So, we obtain a first region of pixels which surely belong to the safe skin and a second region of pixels which surely belong to the lesion.

To define the contour of the lesion as well as it is possible, it is necessary to perform a local approach, what we call the fine segmentation task. For that, a variable Z is defined from an observation vector Y with 14 components built with the following different colour models for any pixel (i,j):

1. the RGB model (3 parameters)
2. the rgb model (3 parameters)
3. the HIS model (3 parameters)
4. the Faugeras model (5 parameters)

The choice of a colour model is often arbitrary. The modelling we adopt take into account different sources of information. We can consider we have nothing less than a fusion of different sources.

A variable Z is obtained from the projection on the direction linking the two conditionnal means computed in the coarse segmentation step. We compute the means and the covariances of the vector Y under each hypothesis. We replace the mathematical expected values by statistical means under the hypothesis  $H_S$  and under the hypothesis  $H_L$ , that is to say:

$$\left\{ \begin{aligned} E(Y/H_S) \approx m_S &= \frac{1}{Card_S} \sum_{(i,j) \in G_S} Y(i,j) \\ Cov(Y/H_S) \approx \Gamma_S &= \frac{1}{Card_S - 1} \sum_{(i,j) \in G_S} (Y(i,j) - m_S)^* (Y(i,j) - m_S) \end{aligned} \right. \quad (2)$$

and:

$$\left\{ \begin{aligned} E(Y/H_L) \approx m_L &= \frac{1}{Card_L} \sum_{(i,j) \in G_L} Y(i,j) \\ Cov(Y/H_L) \approx \Gamma_L &= \frac{1}{Card_L - 1} \sum_{(i,j) \in G_L} (Y(i,j) - m_L)^* (Y(i,j) - m_L) \end{aligned} \right. \quad (3)$$

We consider that Y follows a Gaussian distribution such as:

$$f(Y) = \frac{1}{(2\pi)^{n/2} (\det \Gamma)^{1/2}} \exp \left[ -\frac{1}{2} (Y - m)^* \Gamma^{-1} (Y - m) \right] \quad (4)$$

The logarithm of the likelihood  $\frac{f(Y/H_S)}{f(Y/H_L)}$  gives:

$$\ln \frac{f(Y/H_S)}{f(Y/H_L)} = \frac{1}{2} \ln \left[ \frac{d\det \Gamma_L}{d\det \Gamma_S} \right] - \left\{ (Y - m_S)^* \Gamma_S^{-1} (Y - m_S) - (Y - m_L)^* \Gamma_L^{-1} (Y - m_L) \right\} \quad (5)$$

Let U be the n-dimensional unit vector giving the direction of the centres of gravity such as:

$$U = \frac{\hat{m}_S m_L}{\|m_S m_L\|} \quad (6)$$

Let be T the estimator to choose between the two hypotheses  $H_S$  and  $H_L$ . It is based on the computation of the scalar product  $(\cdot, \cdot)$  of the vector Y of observation and the direction of the centres of gravity, that is to say:

$$T(Y(i,j)) = \sum_{i=1}^n U_i Y_i(i,j) = (U, Y(i,j)) \quad (7)$$

Under the hypothesis  $H_S$ , T is a Gaussian estimator with the mean :

$$E(T(Y)/H_S) \approx \sum_{i=1}^n U_i m_{S_i} \quad (8)$$

and with the variance :

$$V(T(Y)/H_S) \approx U^* \Gamma_S U \quad (9)$$

Under the hypothesis  $H_L$ , we obtain the same formula where L stands for S.

If we compute the projection of the observations on the direction of the conditional means of the two classes (class  $C_L$  of pixels belonging to the lesion and class  $C_S$  of pixels belonging to the safe skin) we define a cut-off direction  $U$ . In the sequel, we consider centered and reduced variables and define the normalized observation vector  $Y'$ .

For any pixel in the image, we define a variable  $Z$  such as:

$$Z = P_U(Y') = \langle Y', U \rangle = Y'_1 u_1 + \dots + Y'_n u_n \tag{10}$$

Because of the geometrical structure of the lesions, we have developed a radial search for the contour localization using the variable  $Z$  (see the Figure 4) for  $m$  orientations of angle  $\theta_m$ . The aim is to locate the contour point in each direction  $\vec{d}_{\theta_m}$  of angle  $\theta_m$ .

Let us note that the separation between the lesion and the safe skin can be inside a blurred zone.

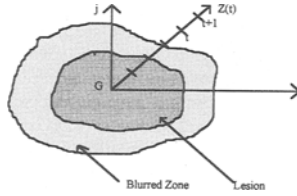


Fig. 3. Radial search principle

Let  $G_L$  and  $G_S$  be respectively the center of gravity of the lesion and the center of gravity of the safe skin region. In each region, we can consider that the variable  $Z$  varies around a mean value given as a straight line. The two defined straight lines (one per region) are quite parallel.

Let us consider the following quadratic cost function:

$$\Phi = \sum_{i \in S} (Z(t) - \hat{Z}_S(t))^2 + \sum_{i \in L} (Z(t) - \hat{Z}_L(t))^2 \tag{11}$$

where  $\bar{Z}_S$  is the expectation of the variable  $Z$  in the safe skin and  $\bar{Z}_L$  the expectation of the variable  $Z$  in the lesion.

So, we compute the estimation  $\hat{a}$  of the slope  $a$  and we have then the equations of the two straight lines:

$$\begin{cases} \hat{Z}_S(t) = \hat{a}(t - \bar{t}_S) + \bar{Z}_S \\ \hat{Z}_L(t) = \hat{a}(t - \bar{t}_L) + \bar{Z}_L \end{cases} \tag{12}$$

The optimal separation between the lesion and the safe skin is obtained by defining a straight line of separation at the middle of the two previous lines. The equation of this straight line is:

$$\hat{Z}_{Sep}(t) = \hat{a}(t) - \frac{1}{2}\hat{a}(\bar{t}_L + \bar{t}_S) + \frac{1}{2}(\bar{Z}_L + \bar{Z}_S) \tag{13}$$

The edge point is then defined as the first point cutting the separation line when one runs along this line from the safe skin side towards the lesion side. We give in figure 4, the behaviour of the variable  $Z$  according to a given direction  $\theta$  and the different straight lines we have obtained. To illustrate the segmentation process, we give in figure 5 the results of the edge location obtained on some lesions.

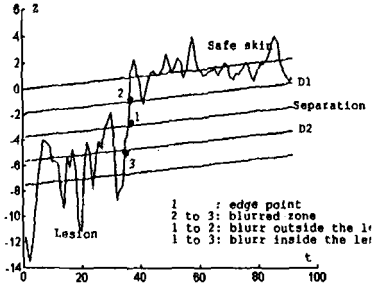


Fig. 4. Edge point location

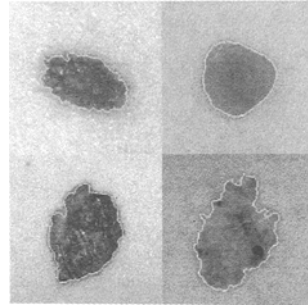


Fig. 5. Contours of different lesions

### 4. Classification

For the classification, we propose 12 primitives - geometric as well as photometric - which are robust and relevant.

The different parameters are:

1. The compactness of the shape.
2. The expanse.
3. The mean deviation from the gravity center.
4. The regularity of the contour.
5. The lengthening
6. The blurred surface quantization.
7. The homogeneity of R.
8. The symmetry of R.
9. The homogeneity of G/R.
10. The symmetry of G/R.
11. The deviation between the mean r on the safe skin and the mean of r on the lesion.
12. The deviation between the mean b on the safe skin and the mean of b on the lesion.

The homogeneity is defined as the sum of the transitions clear zone/dark zone and dark zone/clear zone of C (C=R or C=G/R) when the lesion is described horizontally and vertically. The dark zone is defined as the third darker area of C in the lesion and the clear zone is defined as the third clearer area of C in the lesion.

The colour symmetry is given by:  $s_c = \frac{|\overline{d_D} - \overline{d_C}|}{\overline{d_G}}$  where  $\overline{d_D}$  is the mean distance of the darker pixels to the

gravity center and  $\overline{d_C}$  is the mean distance of the clearer pixels to the gravity center.  $\overline{d_G}$  is the mean distance to the gravity center. Let us note that the homogeneity and the symmetry are indexes of polychromy.

These parameters or primitives derives from clinical features or are original features. It is possible to increase the number of primitives but only robust and efficient parameters are needed in a classification step.

Let be L the whole set of lesions built with the set  $L_B$  of benign lesions and the set  $L_M$  of malignant lesions. We consider the benign lesions as the reference group for which the mean and the standard deviation are computed for each of the  $q$  primitives, with  $q=12$ . So, for a lesion  $l_i$  in  $L_B$ , we define a vector  $X_i$  of dimension  $q$ , such as :

$$X_i = \begin{bmatrix} x_{i1} & x_{i2} & \dots & x_{iq} \end{bmatrix} \tag{14}$$

Let be  $m_{Bj}$  and  $\sigma_{Bj}$  the mean and the standard deviation for the  $j^{th}$  primitive computed on the whole subset  $L_B$  of benign lesions. So :

$$m_{Bj} = \frac{1}{n_B} \sum_{l_i \in L_B} x_{ij} \text{ and } (\sigma_{Bj})^2 = \frac{1}{n_B} \sum_{l_i \in L_B} (x_{ij} - m_{Bj})^2 \tag{15}$$

where  $n_B = \text{Card}L_B$  is the number of lesions in  $L_B$ . Then, we consider the reduced matrix:

$$X' = [Y'_1 \ Y'_2 \ \dots \ Y'_q] \tag{16}$$

built with the centered-reduced vectors defined as it follows:

$$Y'_j = \frac{Y_j - m_{Bj}}{\sigma_{Bj}} \tag{17}$$

A *Principal Components Analysis* (PCA) applied to the results obtained for the benign lesions, gives  $q$  principal axes  $e_j$  of dimension  $q$  which allow to build  $q$  new uncorrelated primitives. We define then  $q$  principal components  $W_j$ , on the whole set of all the lesions.

$$W_j = X' \cdot e_j \tag{18}$$

These  $q$  vectors allow to build the matrix  $W$  which consists of the results obtained on the whole set  $L$  of the lesions:

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1q} \\ w_{21} & w_{22} & \dots & w_{2q} \\ \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{nq} \end{bmatrix} \tag{19}$$

where the scalar  $w_{ij}$  is the value of the  $j^{\text{th}}$  primitive for the  $i^{\text{th}}$  lesion. We want to distinguish the malignant lesions from the benign lesions by the use of the new  $q$  primitives  $e_j$ . A malignant lesion is defined as a lesion for which one or several primitives are far from those obtained on the benign lesions.

Let be  $\text{Hyp}$  the matrix defined by  $nxq$  elements  $\text{Hyp}_{ij}$  as it follows:

$$\text{Hyp}_{ij} = \begin{cases} 1 & \text{if } c_j > m_B(e_j) + k\sigma_B(e_j) \\ -1 & \text{if } c_j < m_B(e_j) - k\sigma_B(e_j) \\ 0 & \text{else} \end{cases} \tag{20}$$

with  $k \in \mathbb{N}$ .

If  $\text{Hyp}_{ij}=1$ , then the lesion  $l_i$  is hyper for the primitive  $j$ . If  $\text{Hyp}_{ij}=-1$ , then the lesion  $l_i$  is hypo for the primitive  $j$  and if  $\text{Hyp}_{ij}=0$ , then the lesion  $l_i$  is indifferent for the primitive  $j$ .

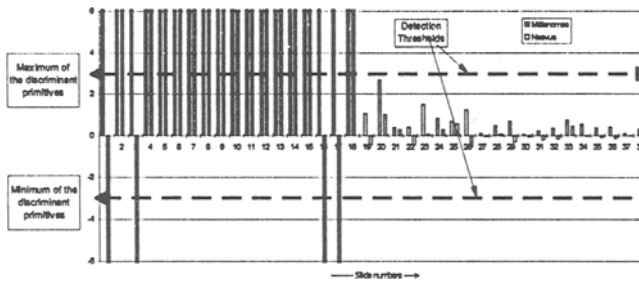
The parameter  $k$  gives the authorized deviation from the mean. In the Gaussian case, it gives the error probability. Taking a weak value for  $k$  leads to an important rate of false alarm. Unlikely, a big value for  $k$  can induce no detection. In practice,  $k=3$  (Probability of false alarm is 0,27%) is a well-suited value.

The more relevant primitives to distinguish the melanoma from benign lesions are those which detect the more numerous malignant lesions and which detect the lesions that others do not detect. The selection of these primitives is based on the principle of overlap [12]. A primitive  $P_i$  overlaps a primitive  $P_j$  when lesions which are considered Hyper or Hypo by  $P_i$  are considered Hyper or Hypo by  $P_j$  too. A primitive overlaps another one if it gives at least the same information. At the end, there are  $d$  primitives among  $q$  which are kept for classification.

### 5. Results

We have applied the classification process on a set of 38 lesions: 20 benign lesions (naevi) and 18 malignant lesions (melanoma). If we take the max or min values of the  $d$  primitives, it appears that the two populations of lesions have a quite different behaviour (see Fig. 6). The two thresholds  $T_{\text{low}}=-3$  and  $T_{\text{high}}=+3$ , allow to detect the melanoma.

**Separation Melanoma/Naevus with primitives Hyper-Hypo**  
 (combination of exponential functions of the geometric and colorimetric primitives)



**Fig. 6. Melanoma detection with the method Hyper-Hypo**

## 6. Conclusion

The method we have developed does not depend on a standard of acquisition conditions. The robustness of the segmentation and the robustness of the primitives used for the classification seem to be real. Furthermore, the feasibility in benign and malignant lesion discrimination are quite possible according to our first results (100% of melanoma detected ; 0% of non-detection or false alarm). However, we have engaged with dermatologists a prospective study in the aim to increase our data base, in particular to increase the number of melanoma. So, we have engaged a test on about 300 slides of benign and malignant lesions.

## Acknowledgments

We thank Professor P. Joly and Professor Ph. Lauret of Charles Nicolle's Hospital for their expert opinion.

## References

- Friedman, R.J. : Early Detection of Melanoma: the Role of Physician Examination and Self Examination of the Skin. *CA*. 35 (1985) 130-151.
- Golston, J.E., Moss, R.H., Stoecker, V. : Boundary Detection in Skin Tumor Images: an Overall Approach and a Radial Search Algorithm. *Pattern Recognition*. 23 (1990) 1235-1247.
- Cascinelli, N., Ferrario, M., Bufalino, R. : Results obtained by using a Computerized Image Analysis System designed as an Aid in Diagnosis of Cutaneous Melanoma. *Melanoma Res*. 2 (1992) 163-170.
- Umbaugh, S.E., Moss, R.H., Stoecker, V. : An automatic color segmentation algorithm with application to identification of skin tumor borders. *Comp. Med. Imag. Graph*. 16 (1992) 227-236.
- Scott, E. : Automatic Color Segmentation Algorithms with Application to Skin Tumor Feature Identification. *IEEE Engineering in Medicine and Biology*. (1993) 75-82.
- Schindewolf, T., Stolz, W., Albert, R., Abmayr, W., Harms, H. : Classification of melanocytic lesions with color and texture analysis using digital image processing. *Anal. Quant. Cytol. Histol*, 15 (1993) 1-11.
- Schindewolf T., Schiffner, R., Stolz, et al. : Evaluation of different image acquisition techniques for computer vision system in the diagnosis of malignant melanoma. *J. Am. Acad. Dermatol*. 31 -1 (1994) 33-41.
- Hall, P.N., Claridge, E., Morris Smith, J.D. : Computer screening for early detection of melanoma - is there a future? *British Jnl of Derm*. 132 (1995) 325-338.
- Colot O., Joly, P., Taouil, K., et al.: Analysis by means of image processing of benign and malignant melanocytic lesions. *European Journal of Dermatology*. 5 (1995) 441.
- de Bruoc D., Taouil, K., Colot, O., et al. : Segmentation d'images et extraction de contours pour l'analyse de lésions dermatologiques. *Proc. Of the 15th Colloque GRETSI, Juan-les-Pins, France (1995)* 1205-1208.
- Kapur, J.N. : A new method for gray-level picture thresholding using entropy of the histogram. *CVGIP* 29 (1985) 273-285.
- Hopcroft, J.E., Ullman, J.D. : *Introduction to automata theory, languages and computation*. Addison Wesley (1979).