Master of science in ICT for smart societies

Report on ICT for Health Laboratory N°2

Moles and clustering

Author:
Bruno Valente

a.a. 2018-2019

1 Introduction

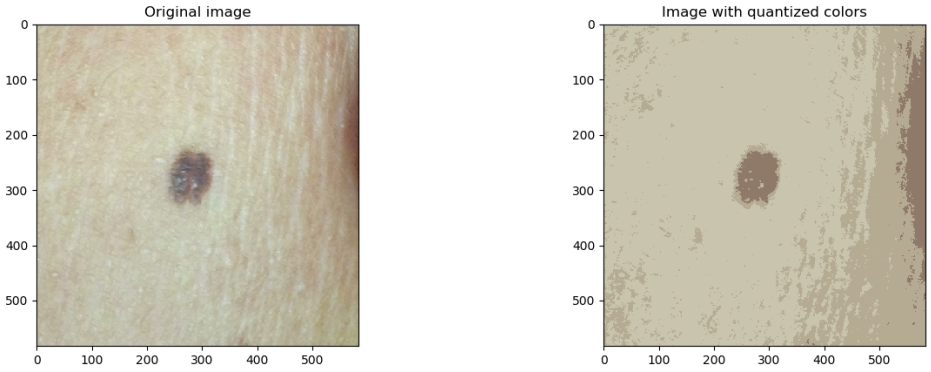
Moles are very common skin lesions caused by a proliferation of melanocytes that grow in clusters instead of being spread throughout the skin.
Since they could develop into melanomas (malignant tumors), it is important to check moles periodically. The features considered by medical doctors in order to detect melanomas are the following five: **A**symmetry, **B**order, **C**olor, **D**iameter, **E**volution. It could be helpful for dermatologists to be provided with ICT tools in order to support them during the analysis of a mole.

Our activities during this project will be focused on the **border** of moles.
The objective of this laboratory, indeed, is to analyse moles from pictures, in order to investigate the correlation between the indentation of the border of the mole and its nature. Other researchers will analyse the other four features. The dataset consists of 54 pictures of moles analysed and classified by medical doctors: 11 of them have been classified as moles which have a low probability of being melanomas; 16 as medium risk moles and the remaining 27 as melanomas.

2 Steps to find the border

2.1 Color quantization and clustering

First of all, it is convenient to simplify the image which has to be analysed. In order to do this, the colors of the image have been quantized using the **K-means** algorithm. The number of the clusters used in the algorithm is 3 for most of the moles. Indeed, that number of clusters is enough for quantizing good quality pictures. For less quality pictures (for example pictures with reflections, shadows, etc.) it would be convenient to increase the number of clusters. Figure 2.1a and Figure 2.1b show an example of original image and the same image with quantized colors, respectively.



(a) Original image.

(b) Image with quantized colors.

Figure 2.1: Original image and quantized image (*low_ risk_ 4*).

2.2 Adjusting the picture

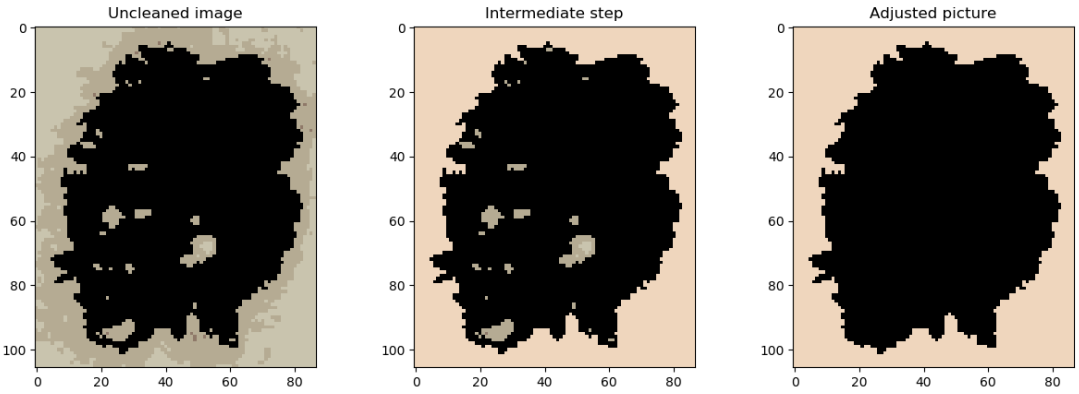
Once the image has been quantized, it could be necessary to make some adjustments. These are really helpful in order to minimize the errors when the algorithm to find the contour of the mole will be performed. The steps that ease the achieving of our purpose are the following:

1. The user is asked to insert the **coordinates** of an internal point of the mole, looking at the picture;

2. Starting from this point, the mole is filled with *black* by using the **flood fill**[1] algorithm;

3. Once the mole has been filled, we are able to localize exactly the pixels belonging to it. Therefore we select a **subset** of the picture including the mole and just a little frame around it;

4. As can be seen in Figure 2.2a, picture imperfections (reflections in that case) could generate **unwanted spots** inside and outside the mole which surely interfere in border measurement. So at this step our aim is to clean the picture removing that spots. To complete this task, we first fill with *pink* the outside of the mole (all the pixels that are not *black*). At this point, the picture has three different kind of pixel (as can be seen in the example in Figure 2.2b): *black* pixels are the ones of the mole; *pink* pixels represent the skin around the mole; all the other pixels are the unwanted spots inside the mole. Now it is simple to detect the pixels that are neither *black* nor *pink* and change them to *black*;



(a) Uncleaned picture. (b) Intermediate step. (c) Final picture.

Figure 2.2: Steps to have a clean picture (*low_risk_4*).

[1] **Flood fill** is a recursive algorithm which performs a DFS (depth-first search), considering the pixels of an image as nodes of a graph. It starts from a seed pixel and terminates when a boundary is found. In other words, it performs what the bucket tool of every graphic editor software does.

2.3 Finding the border

Flood fill algorithm is again the basis for the finding of the border. In fact, it is enough to perform a modified version of the algorithm on the cleaned picture. The mole is filled with *brown* starting from the center of the picture and, when a terminal condition is met (a boundary pixel is reached), the pixels are set to *red*. Figure 2.3 shows an example of mole with its contour in *red*.

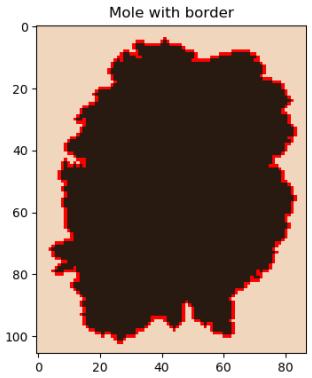


Figure 2.3: Mole with the border (*low_ risk_ 4*).

3 Conclusion

In conclusion, we evaluate the **ratio** between the perimeter of the mole and the perimeter of the circle with the same area. It is simple to find these measurements, because we simply have to count the *red* pixels in the image to evaluate the perimeter of the mole and the *brown* ones to find its area. This ratio gives us information on the **indentation** of the contour.
Figure 3.1 provides histograms of the ratios grouped by category. All the **low risk** moles have a ratio smaller than 1.5; the **medium risk** moles are spread between 1 and 2; most of the **melanomas** have a ratio in the range 1.5 - 2. Some of the analysed melanomas could be considered like *false negatives* because they have a small ratio (i.e. almost circular, regular perimeter). As said in the Introduction, the laboratory is focused only on the border, so this is far from being an accurate classifier.
In Tables 1, 2 and 3 it is possible to notice the evaluated ratios.

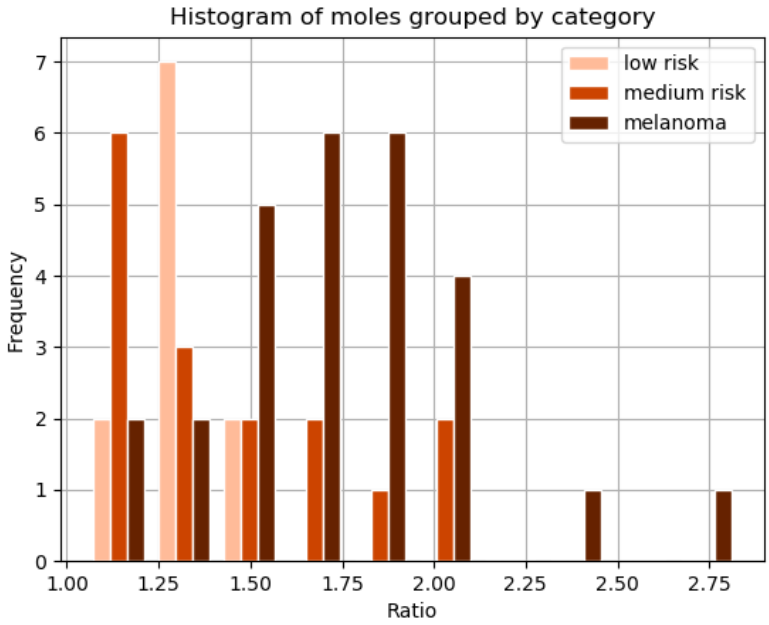


Figure 3.1: Histogram of ratios.

Moles	Ratios
low_risk_1	1.44351
low_risk_2	1.29246
low_risk_3	1.36901
low_risk_4	1.35257
low_risk_5	1.31065
low_risk_6	1.33530
low_risk_7	1.57704
low_risk_8	1.36789
low_risk_9	1.05215
low_risk_10	1.23164
low_risk_11	1.07659

Table 1: Ratios for low risk moles.

Moles	Ratios
medium_risk_1	1.11807
medium_risk_2	1.38742
medium_risk_3	1.07193
medium_risk_4	1.09818
medium_risk_5	1.97987
medium_risk_6	1.69232
medium_risk_7	1.46715
medium_risk_8	1.34402
medium_risk_9	1.97639
medium_risk_10	1.68401
medium_risk_11	1.78352
medium_risk_12	1.14445
medium_risk_13	1.27709
medium_risk_14	1.21479
medium_risk_15	1.22101
medium_risk_16	1.51704

Table 2: Ratios for medium risk moles.

Moles	Ratios
melanoma_1	1.37967
melanoma_2	1.58796
melanoma_3	1.72491
melanoma_4	1.76534
melanoma_5	1.79346
melanoma_6	2.09659
melanoma_7	1.62915
melanoma_8	1.95535
melanoma_9	1.84875
melanoma_10	1.75937
melanoma_11	1.78657
melanoma_12	1.43748
melanoma_13	1.30476
melanoma_14	1.43348
melanoma_15	1.76794
melanoma_16	1.91135
melanoma_17	2.40208
melanoma_18	1.06827
melanoma_19	1.62839
melanoma_20	1.49893
melanoma_21	2.05777
melanoma_22	1.53109
melanoma_23	2.83045
melanoma_24	1.73882
melanoma_25	1.21758
melanoma_26	1.57519
melanoma_27	2.02589

Table 3: Ratios for melanomas.