ICT for Health report 2

Chiara Lanza s277750

Academic Year 2019/20

Introduction

Nowadays telemedicine is necessary and helpful for different reasons.
We can observe that is required teledermatology in particular, because for example more and more people pretend to have a youthful, healthy and attractive skin, which needs the figure of dermatologist.
On the other hand, dermatologists are decreasing in number.
Of course the medical figure is not replaceable by technology's instruments, but they can help reducing the time in diagnosis.

Moles

What we are going to implement is a tool for the medical doctors which can help them to analyze moles.
First of all it is necessary to understand how to recognize a melanoma from a normal mole.
The ABCD rule summerizes the research.
It consideres five features:

- A: Asymmetrical shape. Melanomas are asymmetric moles
- B: Borders. Melanomas are irregular
- C: Color. Melanomas have multiple colors
- D: Diameter. Melanomas have larger ones
- E: Evolving. Melanomas change over time in size, shape and color

Our tool will analyze just symmetry and border getting two features, which will help the classification of the moles together with the other features that are not included in our task.
The dataset is composed by 58 photos of moles: 13 low risk, 16 medium risk and 27 melanomas.

Algorithms

Black and white

The first thing to do is to apply the *k-means* algorithm of the *sklearn.cluster* library.
In order to instantiate the K-means object, the number of cluster is needed as input. The choice of the number of clusters is therefore very important.
The doctor can choose between two versions of the tool.
In the automatic version the 3 main clusters are used by defaut: the darkest one is set to black, the other two to white.
Some moles are not well isolated with this choice because there are more shadows or hairs.
So in the more complex version, the tool asks to the doctor the most accurate version of the clustering. In particular there are three versions of the image which are made by 3, 5 (two of them setted black) and 9 (with three black clusters) main clusters.
The choice is recommended for 10 moles ('low_risk_10.jpg','medium_risk_1.jpg', 'melanoma_5.jpg', 'melanoma_8.jpg', 'melanoma_9.jpg', 'melanoma_10.jpg', 'melanoma_11.jpg', 'melanoma_17.jpg', 'melanoma_21.jpg', 'melanoma_27.jpg') but not indispensable for all of them. The algorithm find the darkest cluster as the smallest number obtained by summing the three values of the RGB triplet.
Every time that a pixel is colored black, a black pixel counter is increased. This parameter is necessary for the next function.
In the figure 1 there is an example of the algorithm with the choice option, where the doctor has to select by command line the most accurate version of the clustering.

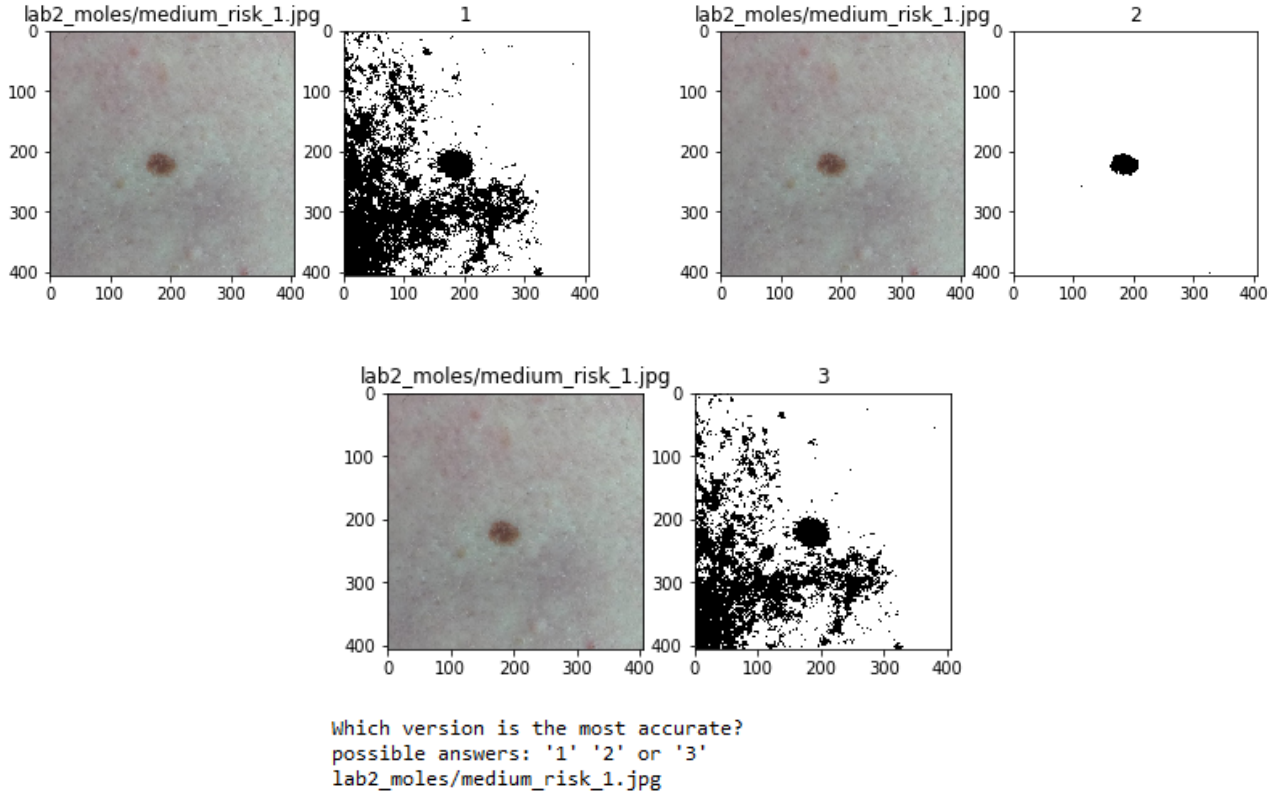


Figure 1: Chioce of the custering: example

Cleaning

The next step is to clean the black and white image from the remaining hairs and shadows.
A first function is applied to the image: resize.

- It takes as input *percB* the percentage of the black pixels in the image and the black and white image.
- The image is therefore cut directly proportional to the percentage of black using a parameter $met = \frac{1-percB}{\eta}$, where η is a proportional value explained soon later.
  The image is 'cutted' for coordinates between $[0, shape*met]$ and between $[shape*(1-met), shape]$ for both x and y, where *shape* is the shape of the image for x and then y.
  An example of what the algorithm does is in figure 2, where the yellow part will be all white in the real algorithm, but it is colored in this case in order to make visible what we are doing.
- The output is the new image

Figure 2: Resize a mole

The function is based on our dataset using the trial and error approach for estimation of the proportional value η and in general assuming that the moles do not occupy the whole photo (which however can happen) and that they are centered.
Another simple function is then applied. It returns as output a matrix with the coordinates of the mole so far isolated and two lists of the x and y coordinates. They are part of the input of the next function: cleanIm.
Two algorithms are implemented in it.

- At first it deletes the black pixels which are statistical outliers, assuming that they are distributed in a Gaussian way.
  It compute the z vaule $z = \frac{(x-\mu)}{\sigma}$ for the coordinates of x and y of the mole detected.
  The z value indicates how many standard deviations the coordinates are away from the means.
  If the z value of the x and y for a determinate point is bigger than a *statistical significance* α, the point is probably noise (not part of the mole) and so it is deleted.
- Then with the same idea of the statistical property of the black pixel distribution, whenever a black pixel with high probability of being part of the inner mole is near a white pixel, this last one is setted to black.

All z values used for the Gaussian cleaning are the result of a trial and error approach.

Perimeter

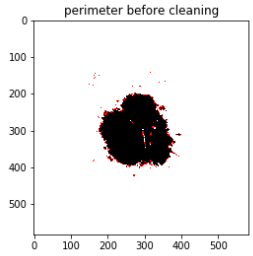
This function is not just the perimeter detection because the mole is not yet fully defined.
In some images there are holes due to the flash for example, which are noisy for the perimeter, the area and the symmetry too. They are deleted in this function because it uses a first perimeter estimation in order to reduce unnecessary cycles.

1. The perimeter at first is calculated in a loop where it's checked if a black pixel is surrounded by at least one white pixel.
   For each point $x_{i,j}$, it consider $x_{i+k,j+m}$, with k and m equal to -1 and 1 alternately.
   If at least one of $x_{i+k,j+m}$ is white, the point $x_{i,j}$ is part of the perimeter.

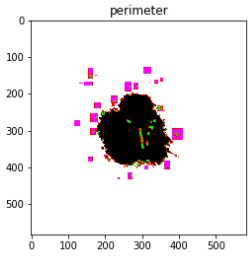
2. This list of pixels is used then to detect holes in the mole.

   - Starting by those pixels, a square/ rectangle is created around them one by one.
   - The polygon is enlarged by one pixel for each direction until it is at the edge of the mole so far detected (the coordinates of the ends are known) or the side in that direction is made up of all black pixels. If all four sides are made up of only black pixels, we are inside the mole so the whole interior is colored black.

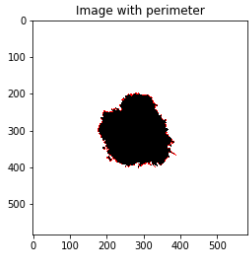
The same algorithm described in point 2 is applied in the opposite way in order to delete outliers that are still presents. There is an additional check for the dimension of the rectangle that should be 'deleted', because if it is too large (more than 30 pixel in a side) probably is part of the mole.
Now the perimeter is calculated with the algorithm in point 1.
An example with one of the most critical mole of the first perimeter detection is in 3a, and after the holes cleaning and pixels out of the image is 3c.
In particular in the 3b image the green parts are the rectangles that will be setted to black, the pink ones will be delete. The original photo was 3d and next to it the perimeter is 3e.



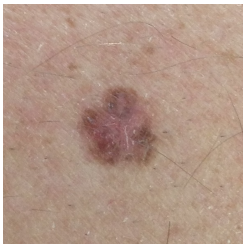
(a) Before the cleaning



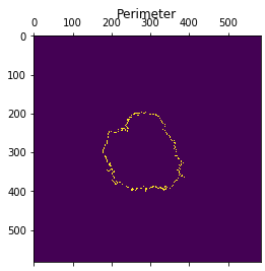
(b) Cleaning explanation



(c) Final perimeter



(d) Original photo



(e) Perimeter

Figure 3: Perimeter

Ellipse

In order to find the axes of symmetry, the geometric moments are studied because they give information about the distribution of the pixels inside the image (the mole in our case).
The second order moments are used to determine the main axes of inertia of the object in the image.
They define an approximation of the object called image ellipse, which is an ellipse with the same area, orientation, eccentricity of the mole and it is centered in correspondence with its centroid.
In general the moments are calculated as $m_{(\mathrm{pq})}=\sum_x \sum_y x^p y^q f(x,y)$

In particular:

- m_{00} is the area of the mole
- m_{10} and m_{01} are useful to find the coordinate of the centroid $x_\mathrm{c} = \frac{m_{10}}{m_{00}}$ and $y_\mathrm{c} = \frac{m_{01}}{m_{00}}$
- the central moments invariant to the translation are $\mu_{pq}=\sum_x \sum_y (x - x_\mathrm{c})^p * (y - y_\mathrm{c})^q * f(x,y)$
- the parameters for the image ellipse are :
  $\alpha = \left(\frac{2[\mu_{20}+\mu_{02}+\sqrt{(\mu_{20}-\mu 02)^2+4\mu_{11}}]}{\mu_{00}}\right)^{1/2}$ $\beta = \left(\frac{2[\mu_{20}+\mu_{02}-\sqrt{(\mu_{20}-\mu 02)^2+4\mu_{11}}]}{\mu_{00}}\right)^{1/2}$ the two semi-axes
  $\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right)$ angle of the semi-axes of the ellipse with respect to the main axes

Once calculated them in this simple way, the function returns these parameters in order to draw on the image the ellipse as is shown in figure 4 and then use the theta to turn the mole.

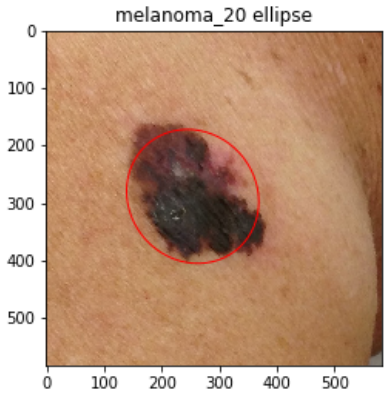


Figure 4: Ellipse

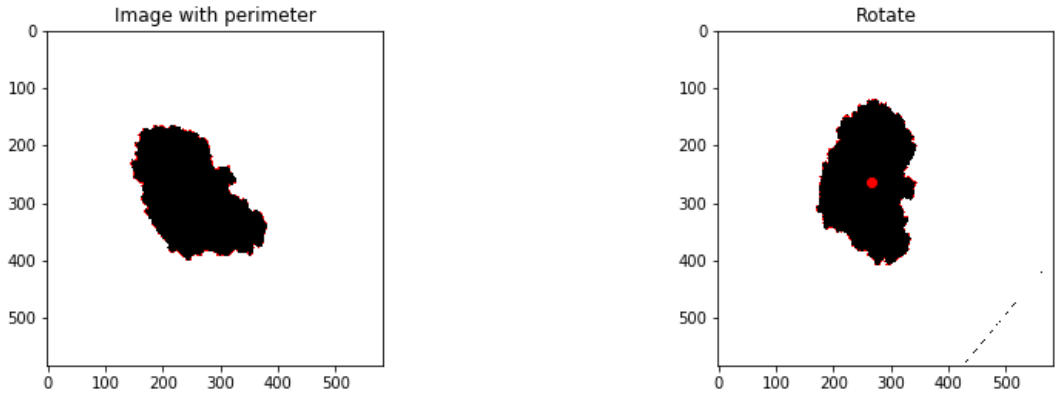
Turn

This function takes as input the extremes of the mole, the image processed so far, the coordinates of the centroid of the mole and the angle θ.
There is a sub function which is called for each pixel.
In particular its actions are:

- Translate the coordinates so that the center of rotation coincides with the origin, by subtracting the coordinates of the center of the image.
- Then the rotation is applied with the rotation matrix $R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$
- At the end there is the reverse of the translation process by summing the coordinates of the center.
- Check if the new coordinates are out of the dimension of the image

In matrix form, the process is $\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_1 - x_0 \\ y_1 - y_0 \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$

Now we have to solve the so called *aliasing* problem.
The pixel boundaries are at quantized locations, sin and cos return real numbers, so we have to round them, and the result is to miss some pixels making holes in the turned image.
The workaround is to set the white pixels surrounded by other black pixels to black.
Analyzing the same mole of the previous figures, the image before applying the turn function is 5a, and the turned mole is 5b.



(a) Before the function

(b) Rotated mole

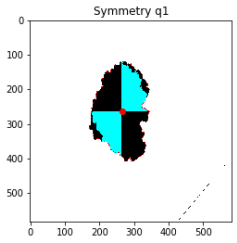
Figure 5: Turn function

Symmetry

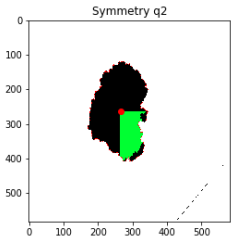
In order to evaluate the symmetry of the mole, it is divided ideally in the four quadrants delimited by the axes.
We turn it before, so the axes we are looking for are simply the parallel axes passing through the center (centroid).
Symmetry algorithm:

- First of all it computes the matches between each pixel of a quadrant and the corresponding one with respect to the ideal axis considered.
- Then an index is given by the ratio between the biggest area from the two quadrants and the matched area.
- The single parameter is given by the sum of these 4 index minus 4.
  All of them will be of course bigger than 1 (it choose the numerator so that is bigger then the matched area).

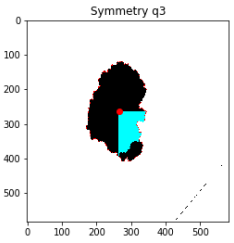
In order to make visible what we are doing, every time a pixel is matched, is colored in a different way.
For the last time, with melanoma_20 in figure 6 , we can see how the algorithm works.



(a) Projection of the quadrant top left on the top right and down left



(b) Projection of the quadrant top right on the down right



(c) Projection of the quadrant down left on the down right

Figure 6: Symmetry

Conclusion

At the end of the run, the two parameters will be wrote in two different files : Ratio and Symmetry.
Some output figures of intermediate functions are saved in the rispective folds.
In the tables 7 there are the two features of the low risk, medium risk and melanoma moles divided by category. The features are the ratio between the real perimeter calculated by the tool and the ideal perimeter of a circle with the same area of the mole, and the symmetry index derived by the idea before explained.
The statistical values are in the last rows.
There is a gray row in table 1, because its mole's picture in the given dataset does not seem a real photo, but it seems a clustered image of another mole present in the dataset. Its values are not good for the estimations. They are outliers because it was difficult to isolate the possible mole. The statistical data without this row are in the last two rows.
The values we found can help the doctor to understand the order of magnitude of them for each category of moles.
Even if they are close each other, the progressiveness of the mean of the ratio and the symmetry index from low risks to melanomas is evident. Obviously they will not replace the doctor judgment, but they will reduce the processing time and the final evaluation if used properly.
These are just two of the five features necessary for the detection of a possible melanoma, therefore the fact that the mean of the symmetry index of the low risk mole is higher than the medium risk one does not mean that moles with less risk of becoming tumor have less symmetry then moles that could become it. They will not be used alone, but a more complex tool is needed.
About precision and accuracy of the tool, if the number of principal centroids chosen are always the same, the results will not change a lot, even if the k-means algorithm does not necessarily find the most optimal configuration.
Of course with a different number of centroids the whole image changes and the extracted features are not the same.
The real data that should have the moles are not given to us, so it is not possible to estimate how much accurate the tool is.
In general, by analyzing the mean values of the features, which are smaller for the low risk moles and then increase for medium risk and melanoma ones, we can say that th results are as we expected.
The standard deviation of the features are not small enough compared to mean values, sometimes they are very close. They depend on the dimension of the moles, how much the functions are accurate and must be taken into account that these will not be the only features used for the classification.

Mole	Ratio	Symmetry
1	1.74	0.47
10	1.29	0.24
11	1.36	0.30
2	1.56	0.28
3	1.64	0.31
3h	1.40	0.31
3s	2.36	2.58
4	1.60	0.76
5	1.65	0.60
6	1.53	0.34
7	1.81	0.61
8	1.67	0.53
9	1.38	0.21
Mean	1.62	0.58
Std	0.28	0.62
Mean 2	1.55	0.41
Std 2	0.16	0.17

Table 1: Low risk

Mole	Ratio	Symmetry
1	1.31	0.35
10	1.78	0.37
11	2.08	0.49
12	1.48	0.31
13	1.51	0.35
14	1.46	0.41
15	1.51	0.35
16	1.72	0.20
2	1.61	0.36
3	1.32	0.24
3h	1.31	0.20
3s	1.45	0.28
4	1.34	0.48
5	2.10	0.39
6	1.99	0.77
7	1.79	0.41
8	1.58	0.30
9	2.02	0.45
Mean	1.63	0.37
Std	0.27	0.13

Table 2: Medium risk

Mole	Ratio	Symmetry
1	1.63	0.40
10	2.36	0.65
11	2.21	0.49
12	1.60	0.35
13	1.61	0.43
14	1.83	0.63
15	2.09	0.96
16	2.15	0.80
17	2.14	0.56
18	1.41	0.35
19	1.89	0.81
2	1.80	0.71
20	2.26	0.54
21	2.34	0.52
22	1.80	0.53
23	5.04	2.71
24	2.82	1.10
25	1.64	0.57
26	1.94	0.93
27	3.17	1.35
3	2.14	0.67
4	2.12	0.50
5	1.81	0.58
6	2.25	1.26
7	1.72	0.37
8	2.07	0.35
9	3,19	0.58
Mean	2.20	0.75
Std	0.88	0.48

Table 3: Melanoma

Figure 7: Features