

PRIM  
Comparison of Segmentation Methods for Skin  
Lesions in Dermoscopy Images

Roman Fenioux

September 2017 - January 2018

## Introduction

Dermoscopy is a technique used by dermatologist to examine skin lesion, in order to distinguish cancerous lesions, especially melanoma, from benign lesions. However, it is a tedious and challenging task for dermatologist because of the visual similarity between the types of lesions and the important variations of shape, color and texture. This lead to a significant intra and interpersonal variability and a diagnosis accuracy of only 60% [7]. An efficient computerized analysis of dermoscopy images represents therefore a huge benefice while providing more reproducibility in the results and quantitative information. Such methods have usually three main steps : segmentation, feature extraction and classification. Although recent methods based on deep neural network merge the feature extraction with the classification, they often still have a distinct segmentation stage.

This project will focus on the segmentation stage, for which we will compare three methods based on different approaches : thresholding [5], region growing [2] and level set active contours [9]. Other methods include clustering [6], edge operators, morphological watershed [13], model-based algorithms [4] and soft computing [14]. Overview of the recent border detection methods has been conducted in 2009 by Celebi and al [1].

## Data

The data used for this project are 42 images from the 2017 ISBI challenge "Skin Lesion Analysis Toward Melanoma Detection" [3]. The sizes range from 538x720 to 2814x2110 pixels for an 8 bit encoding. Segmentation masks were created by an expert clinician, using either a semi-automated process (using a user-provided seed point, a user-tuned flood-fill algorithm, and morphological filtering) or a manual process (from a series of user-provided polyline points).

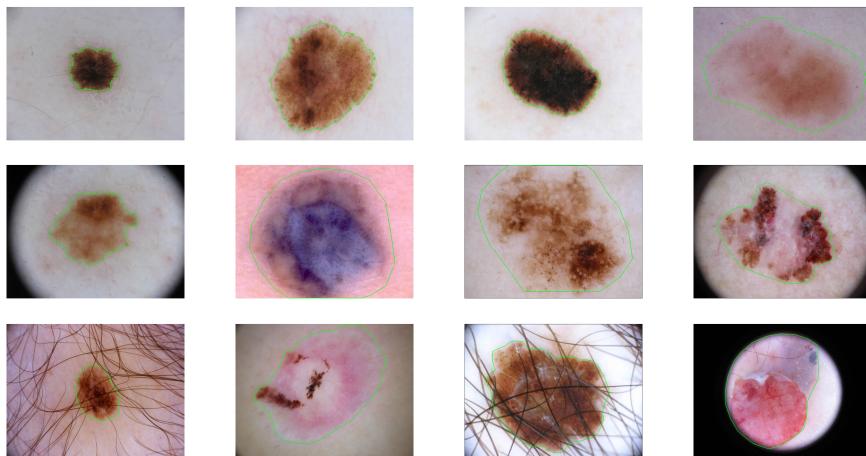


Figure 1: Example of data used with ground truth segmentation (green)

Examples can be seen in figure 1 and we can note the difference between the semi-automated method that closely follows the lesion borders as in the upper

left image and the polyline method, purposed to contain all of the lesion area when border are fuzzy as in the top right image.

## Problems with border detection

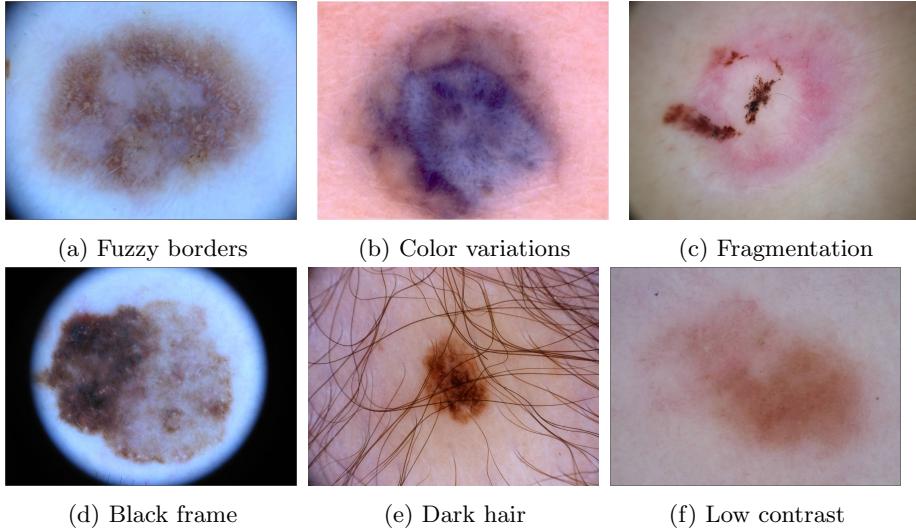


Figure 2: Most common problems encountered in lesion segmentation

The segmentation errors are often caused by two main reasons. The first is a problem of image quality such as the presence artifacts (hair and other objects, black frame around the lesion) or a low contrast. The second is the properties of the lesion itself : fuzzy borders, variation of the color of the lesion or the surrounding skin, and fragmentation of the lesion in multiple nodules. These conditions make the use of simple automated segmentation methods challenging without other processing stages.

## 1 Pre-processing

### 1.1 Conversion to grayscale

**Choosing a color channel** The segmentation performance is influence by the color space used. Although the literature commonly uses the blue channel from RGB color space as in [10], a paper from Garnavi, Celebi and al [5] has shown that for a threshold-based approach the X channel from CIE-XYZ color space could perform better.

To decide which of the blue or the X channel was the best for the ISBI dataset, a comparison of the segmentation performances was conducted. To have a better understanding of the influence of the color channel selection, this comparison also included red and green channels from RGB, luminance from HSV and the arithmetic averaging of the three RGB channels (meanRGB).

Results for all channels are shown in table 1 and with more details for the three best channels in figure 3. Despite one severe failure (please mind that the

Table 1: Color channel influence on thresholding results

channel	R	G	<b>B</b>	X	V	meanRGB
Average dice	0.78	0.87	<b>0.89</b>	0.85	0.78	0.87
Average jaccard	0.65	0.78	<b>0.81</b>	0.75	0.66	0.77

y axis ranges from 0.5 to 1), the blue channel is still the best one in average, followed by the green channel and meanRGB. Therefore, the blue channel was chosen for the conversion from RGB to grayscale image.

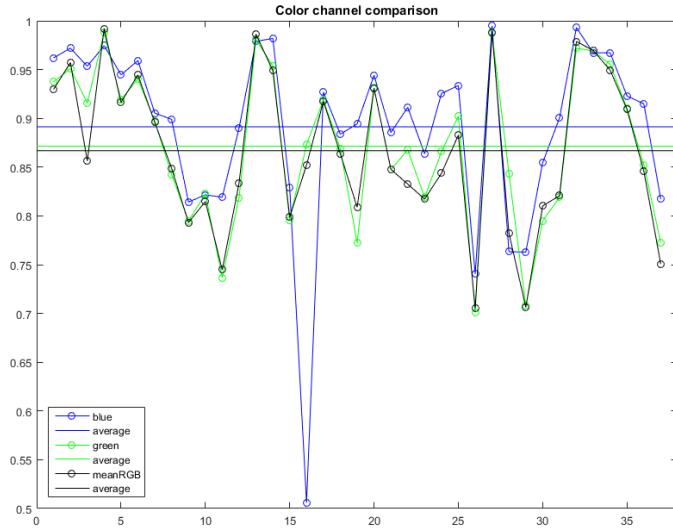


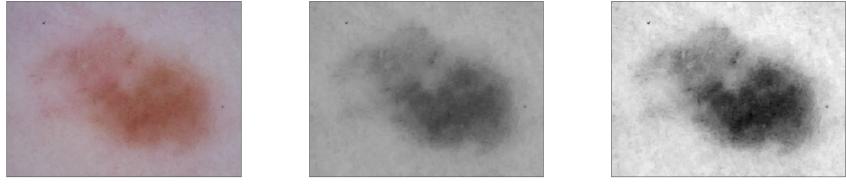
Figure 3: Dice index of the 3 best channels (B, G, and meanRGB)

**Contrast enhancement** We maximize the contrast with a linear contrast transformation by first subtracting the min value and then dividing the obtained image by its max value. This step is done after the removal of black border, so that only the area containing the skin and the lesion is taken into account for this transformation.

## 1.2 Hair removal

Dullrazor [8] algorithm is a standard method for hair removal. It consists in the following steps:

- Locating dark hair with a grayscale morphological closing operation with vertical, horizontal and diagonals structure elements on the three RGB channels. We obtain the hair mask by thresholding the absolute difference with the original image.



(a) Original image      (b) Blue channel selected      (c) Contrast enhancement

Figure 4: Conversion to grayscale

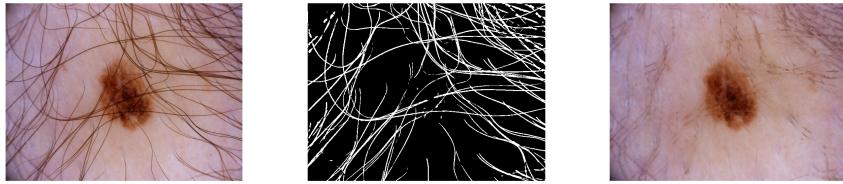


Figure 5: Importance of hair removal for segmentation

2. Denoising the hair mask and interpolating the hair pixels with neighbor non-hair pixels.
3. Removing the remaining hair artifacts with an adapted median filter, only applied on the pixels located in the enlarged hair regions (morphological dilatation of the hair mask).

Example in figure 5.

The drawback of this method is that all parameters are adapted to be efficient on 512x486 pixels images, whereas the dataset used here contains high resolution images of different sizes (from 720x538 to 2814x2110). The images were therefore resized to the smallest resolution to reduce runtime of the segmentation algorithms and the parameters of the hair removal algorithm were adapted to fit this resolution.

### 1.3 Black frame removal

We see in figure 6 how the performance increases with the removal of the black frame. Thresholding methods are the most affected since they are non local : the black corner of the image are the darker areas of the image and will therefore be selected as part of the lesion. One way of solving this issue is to perform a connected component analysis and select the ROI among the different clusters in postprocessing.

However, this approach is not satisfying because firstly the lesion can be connected to the black frame (figure 6 bottom) and secondly it can have multiple nodules so we can't just keep one "best" region according to any criterion. An additional reason to perform this step in preprocessing for the thresholding methods is to avoid any impact on the selection of the optimal threshold towards the black.

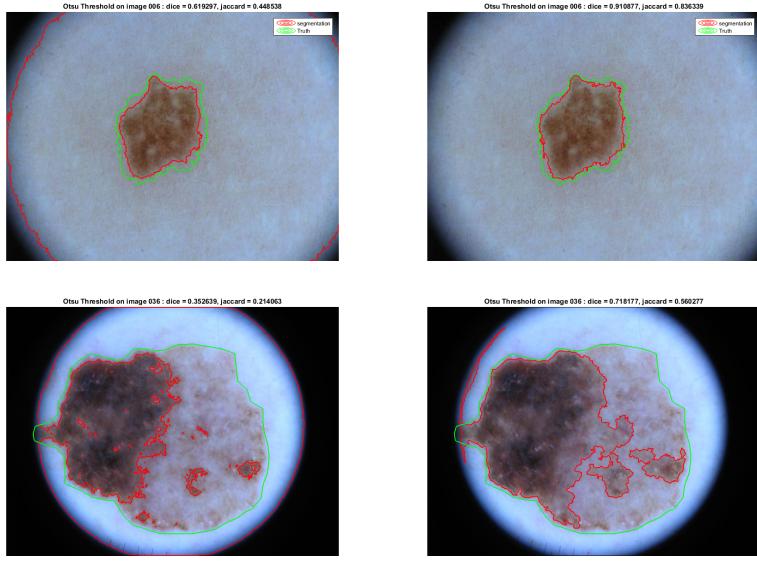


Figure 6: Importance of black frame removal

We can observe this effect on the figure 6 : the ROI expands slightly on the right image compared to the left one. The black frame has been identified before the segmentation process, so that it is not taken into account when choosing the threshold.

## 2 Post-processing:

**Morphological filling** This step consists in filling the holes that can appear inside the ROI with certain methods, such as thresholding or region merging. Our knowledge of the topology of the lesion is that they must be segmented in a unique connected component without hole. This is illustrated in figure 7b.

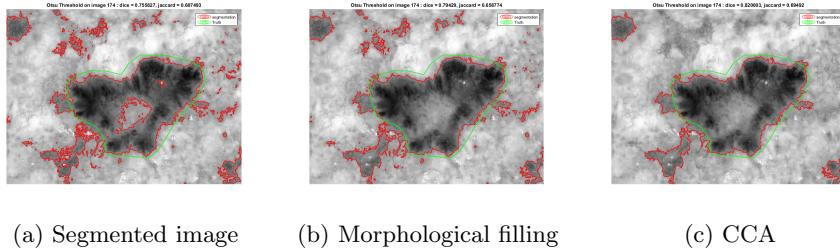


Figure 7: Post processing steps

**Connected Component Analysis** One of the major drawbacks of a thresholding is its non-local approach. Because of this, the segmentation result can be fragmented in several regions. A simple way to solve this issue would be to

keep the biggest region, but it is not always the best choice, since lesion can be fragmented. The ideal result would be to merge the fragment by filling the space in between, but it is hard to distinguish a lesion fragment from a segmentation error caused by artifacts. Therefore, the choice made was to eliminate small regions (with an area more than 10 times smaller than the biggest region). This criterion was developed experimentally by trial and error on an image set distinct from the evaluation set. Result on an image can be seen in figure 7c.

### 3 Adaptative thresholding

#### 3.1 Method

Thresholding is the simplest of the 3 methods reviewed here. It takes advantage of the fact that skin lesions are usually darker than the surrounding skin. Given a grayscale image, we can classify pixels as being part of the region of interest (ROI) or of the background based on their intensities. However, using an absolute threshold for all images is not possible because of the color variability.

Otsu's algorithm [12] gives a way to find an optimal threshold by seeing the histogram as a probability distribution, and choosing the threshold  $k^*$  that maximizes the separability measure  $\eta$  (equation 1) or equivalently the interclass variance  $\sigma_B^2$  because the total variance  $\sigma_T^2$  remains constant for any chosen threshold. So the problem comes down to finding the threshold that maximizes  $\sigma_B^2$  among the  $L$  possible levels of gray as defined in the equation 2.

$$\eta = \frac{\sigma_B^2}{\sigma_T^2} \quad (1)$$

with  $\sigma_B^2 = \mathcal{P}(C_0)\mathcal{P}(C_1)(\mu_1 - \mu_0)^2$  the interclass variance

$$k^* = \underset{1 < k < L}{\operatorname{argmax}} \sigma_B^2(k) \quad (2)$$

#### 3.2 Implementation

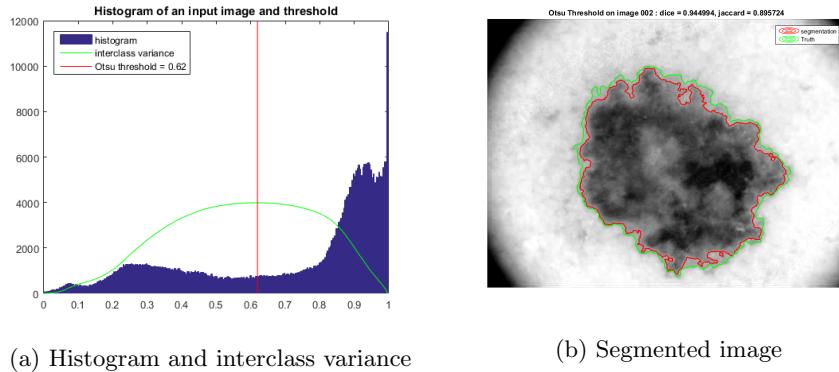


Figure 8: Otsu's threshold

The implementation of this method is straightforward. We build the histogram of the image, and search the best threshold among the 254 possible values. For each threshold value, the histogram is divided into two clusters  $C_0$  and  $C_1$ . We compute the associated inter-class variance using the equation 2 and keep the threshold corresponding to the maximum. The figure 8 shows the segmented image and its histogram (note that the black corners are not included thanks to the black frame removal). We can notice that despite a very flat evolution of the histogram, the chosen threshold can separate the two clusters in the middle which lead to the best separation of the clusters.

However, it has been discussed [1] that human operators tend to put the separation border on the outmost detectable pigmented pixel, whereas this threshold select an intermediary value. This can lead to under-segmentation.

### 3.3 Results

**Metrics** Two metrics are used for the evaluation of the results.

The dice index is calculated of the as shown in the equation 3:

$$d = \frac{|I \cap J|}{1/2(|I| + |J|)} \quad (3)$$

The jaccard index is calculated as shown in the equation 4:

$$j = \frac{|I \cap J|}{|I \cup J|} \quad (4)$$

This two indices are similar but jaccard is slightly more punishing, especially in cases where the segmented region and the ground truth have the same area but are not perfectly overlapped.

**Results without processing** The two top rows of the figure 9 show examples of segmentation results obtained without any pre- or post-processing other than the conversion from RGB to grayscale on the blue channel. As expected, this approach is very sensitive dark hair, black frame and fuzzy borders, but gives very good results on well contrasted lesions with sharp edges. On the entire dataset, the performance is :

Average dice index = 0.75

Average jaccard index = 0.65

**Amelioration with pre- and post-processing** This increases the performances significantly :

Average dice index = 0.84

Average jaccard index = 0.75

The bottom rows of the figure 9 help us to see how the segmentation changes with the additional processing stages.

**Melanoma and nevus** Results on the entire dataset are shown in figure 10, with a separation between melanoma and nevus classes. We can observe a significant difference of performance between images of benign and cancerous lesions : the results are better on images of nevus. An explanation for this effect

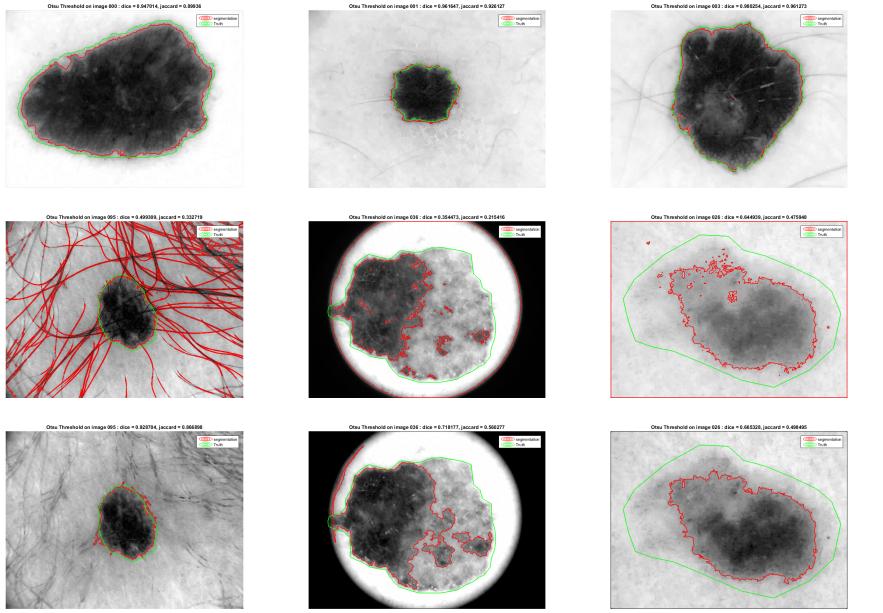


Figure 9: Thresholding segmentation examples

First and second row : no processing - bottom row : with processing

could be the more complex structure of cancerous lesions, that often present more color variation and shape irregularities.

The critical failure of the 19<sup>th</sup> nevus image is caused by a very unusual lesion structure, shown in figure 11. In this specific case, the blue channel is completely blind to the red variation and is not able to correctly detect the true borders.

## 4 Region based segmentation

### 4.1 Method

First, a simple region growing algorithm, but too slow and bad performances so SRM was used

**Statistical Region Merging [11]** This method has been used by Celebi and al [2] to perform a fast and efficient segmentation of skin lesions. This approach consists in merging regions in a particular order, with a statistical criterion.

The image is considered as an observation of a theoretical image made of two regions : the lesion and the skin. We assume an homogeneity property of those regions, so that all pixels in a given region have the same statistical expectation. Each region can be described by a distinct probability distribution.

SRM is based on two concepts : a merging predicate, described in equation 5 and an order two follow. This order is given by a similarity function  $f$ , which ensures that when merging two growing regions, all tests inside the regions have

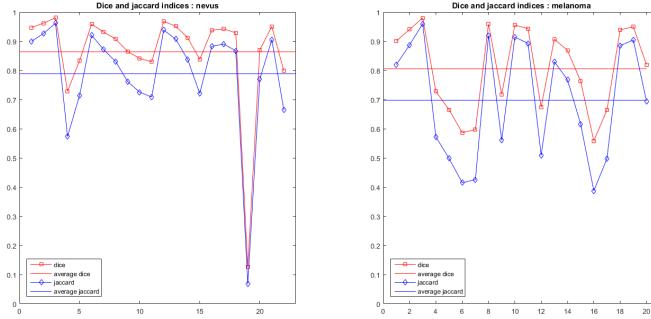


Figure 10: Adaptative thresholding, results for melanoma and nevus

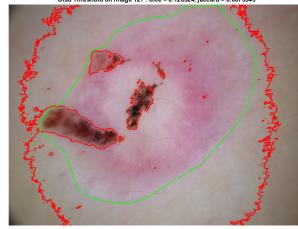


Figure 11: Critical failure for Otsu thresholding

occurred. This results in a huge diminution of the runtime. A possible function is the radiometric difference as defined in equation 6.

$$P(R, R') = \begin{cases} \text{true} & \text{if } \forall a \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}, |\bar{R}'_a - \bar{R}_a| \leq |\sqrt{b^2(R) + b^2(R')}| \\ \text{false} & \text{otherwise} \end{cases} \quad (5)$$

$$\begin{aligned} b(R) &= g \sqrt{\frac{1}{2Q|R|} \ln(6|I|^2 R_{|R|})} \\ f(p, p') &= |p - p'| \end{aligned} \quad (6)$$

This method requires one parameter : Q. It is used to compute the threshold value  $b$  of the merging predicate  $P$ . A high value for Q implies a lower tolerance for merging region. This will result in a more fragmented output but also more homogeneous regions.

**Why SRM?** Compared to traditional region growing, the SRM algorithm has several advantages :

- No initialization required, which can have a heavy impact on the performances and the reproducibility.
- Fast
- Less sensitive to local inhomogeneities
- Better merging order (the more similar region first)

- Can produce more than two region : useful for fragmented lesions but that implies deciding the region to keep in the end.

## 4.2 Implementation

The implementation follows [2] and consists in the following steps :

1. Preprocessing stages (conversion to grayscale, contrast enhancement, black frame detection, hair removal)
2. Adapted median smoothing of size  $n = \text{floor} \left( 5\sqrt{(M/768)(N/512)} \right)$  for better region homogeneities. The size of the mask is adapted to the size of the image given the observation that  $n=5$  is a good choice for images of size 768x512.
3. Pixel pairing in 4-connectivity and sorting by increasing order of  $f(p, p')$
4. Traversing the order once and perform the merging test for any pair of pixels where  $R(p) \neq R(p')$ . Merge if it returns true.
5. User input : selection of a skin patch.
6. Classification of the region between lesion and skin by comparing their mean value to the skin patch.
7. Postprocessing stages

The steps 3 and 4 correspond to the actual SRM algorithm. Since it produces more than two regions, as shown in figure 12b, we use a skin patch to distinguish skin regions from lesion. [2] does it automatically by choosing the skin patch in the corner of the image, but it was not possible with my data since there is some times no location in the image on which we are sure to find skin (the corners are sometimes covered by the black frame). Therefore, we used a input from the user to select a patch of skin.

Then, we compute the mean value of this patch, and we reject as skin all region with a mean value closer than 60 to the skin mean value. The result of the segmentation obtained can be seen in figure 12c.

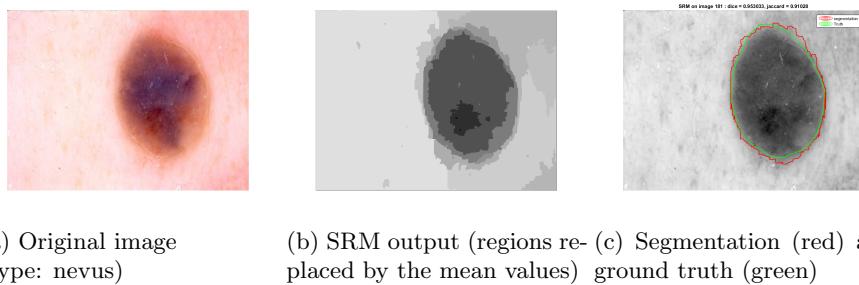


Figure 12: Post processing steps

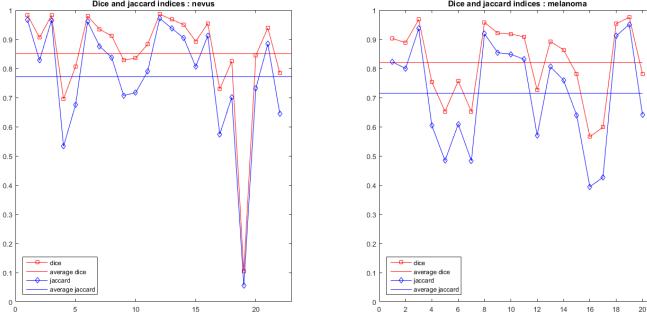


Figure 13: Statistical Region Merging : results on melanoma and nevus

### 4.3 Results

Average dice index = 0.84

Average jaccard index = 0.74

Again, we can observe in figure 13 better performances on nevus images. The nevus image number 19, discussed earlier, is also challenging for this method because it is very fragmented and therefore get separated into different regions.

The weak point of this method is the region selection. Instead of simply comparing the mean value of the skin patch to those of the regions, we perform a better texture analysis by including colors, or variance. Another approach could be to select a lesion patch instead of a skin patch and choose similar region.

## 5 Parametric active contours

### 5.1 Method

The third method studied here takes the approach of active contour, Distance Regularized Level Set Evolution [9]. The contour is described as the intersection of a higher dimension Level Set Function (LSF) with the level 0, as expressed in equation 7.

$$\phi_k(x, y) = 0 \quad (7)$$

The evolution of the curve is then described in 8 by the following partial differential equation, or level set evolution equation :

$$\frac{\partial \phi}{\partial t} = F |\nabla \phi| \quad (8)$$

To maintain a stable evolution and an accurate numerical computation, we need to keep the LSF smooth, not too steep or too flat. This property is satisfied in the signed distance function where  $|\nabla \phi| = 1$  (equivalent to a 45 degree angle of the 3D surface in our case). We can see an illustration of the distance function initialized with 3 points in figure 14

The distance regularization proposed in [9] helps maintaining this property. It uses an energy functional  $\mathcal{E}$  defined in 9

$$\mathcal{E}(\phi) = \mu \mathcal{R}_p(\phi) + \mathcal{E}_{ext}(\phi) \quad (9)$$

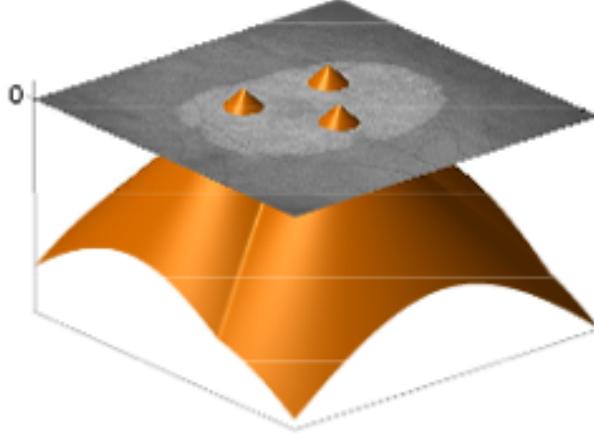


Figure 14: Distance Function

Where  $\mu > 0$ ,  $\mathcal{E}_{ext}$  is the external energy derived from the gradient of the image, and  $\mathcal{R}_p$  is the distance regularization term defined in

$$\mathcal{R}_p(\phi) \triangleq \int_{\Omega} p(|\nabla \phi|) dx \quad (10)$$

where  $p$  is a potential function  $p : [0, \infty) \rightarrow R$

To avoid some side effects, [9] uses a double-well potential  $p = p_2$  that has two minimum points at  $s = 1$  and  $s = 0$ . This function is explicitly defined in 11

$$p_2(s) = \begin{cases} \frac{1}{2\pi} \sin(2\pi s) & \text{if } s \leq 1 \\ \frac{1}{2}(s-1)^2 & \text{if } s \geq 1 \end{cases} \quad (11)$$

Finally, the Distance Regularized Level Set Evolution (DRLSE) can be expressed as in 12. This equation is derived from the variational formulation 9.

$$\frac{\partial \phi}{\partial t} = \mu \cdot \operatorname{div}(d_p(\nabla \phi) \nabla \phi) - \frac{\partial \mathcal{E}_{ext}}{\partial \phi} \quad (12)$$

## 5.2 Implementation

### Discretization

The DRLSE in 12 is implemented with a finite difference scheme where spatial derivatives are replaced by central difference with a fixed step of  $\Delta x = \Delta y = 1$  and the temporal derivative is approximated by the forward difference. We obtain the following equation

$$\phi_{i,j}^{k+1} = \phi_{i,j}^k + \Delta t L(\phi_{i,j}^k), \quad \text{with } k = 0, 1, 2, \dots \quad (13)$$

For the segmentation, the edge indicator is given by the function defined in 14, where  $G_\sigma$  is a gaussian kernel of standard deviation  $\sigma$  to reduce noise. g

takes small values on edges.

$$g \triangleq \frac{1}{1 + |\nabla G_\sigma * I|^2} \quad (14)$$

The energy can then be expressed as 15 :

$$\begin{aligned} \mathcal{E}(\phi) &= \mu \mathcal{R}_p(\phi) + \lambda \mathcal{L}_g(\phi) + \alpha \mathcal{A}_g(\phi) \\ \mathcal{L}_g(\phi) &\triangleq \int_{\omega} g \cdot \delta(\phi) \nabla \phi dx \\ \mathcal{A}_g(\phi) &\triangleq \int_{\omega} g \cdot H(-\phi) dx \end{aligned} \quad (15)$$

$\delta$  is the Dirac function, approximated by the equation 16, so that  $\mathcal{L}_g(\phi)$  computes the integral of  $g$  along the line where  $\phi = 0$ , that is the actual contour used for segmenting the image. This integral is minimized when this contour is located at the boundaries of the lesion, where  $g$  takes its smallest values. This component is weighted by the parameter  $\lambda$ .

$$\delta_\epsilon(x) = \begin{cases} \frac{1}{2x} [1 + \cos(\frac{\pi x}{\epsilon})], & \text{if } |x| \leq \epsilon \\ 0 & \text{if } |x| > \epsilon \end{cases} \quad (16)$$

$H$  is the Heaviside function, approximated by the equation 17, so that  $\mathcal{A}_g(\phi)$  computes a weighted area of the region where  $\phi(x) < 0$ . It is used to force the contour to grow or to shrink by changing the value of  $\alpha$ . Since this paper gives negative values inside the contour and positive values outside, we take positive values of  $\alpha$  so that the contour can shrink, with an initial contour placed outside of the lesion. The presence of  $g$  in this energy term will slow down the shrinking when we arrive at the boundaries of the lesion.

The decision of having a contour initialized outside of the lesion and a shrinking evolution was mainly made to speed up the algorithm : after the initialization, the image was cropped to the bounding box of the initial contour, as we know that the lesion is contained in it.

$$H_\epsilon(x) = \begin{cases} \frac{1}{2} [1 + \frac{x}{\epsilon} + \frac{1}{\pi} \sin(\frac{\pi x}{\epsilon})], & \text{if } |x| \leq \epsilon \\ 1 & \text{if } |x| > \epsilon \\ 0 & \text{if } |x| < -\epsilon \end{cases} \quad (17)$$

### Algorithm

- Initialization of  $\phi$  with  $\phi_0$  provided through user input (polyline).
- Update the LSF :  $\phi_{i,j}^{k+1} = \phi_{i,j}^k + \tau L(\phi_{i,j}^k)$
- Assign values to new pixels
- Test the termination : stop if the zero crossing points stop varying or  $k$  exceed a maximum iteration number, else go to step 2.

## Parameters

The DRLSE algorithm requires several parameters listed below, and proved to be very sensitive the values used. I managed to obtain good results using the following values:

- Timestep:  $\tau = 1$ ,
- Distance regularization:  $mu = 0.2/\tau = 0.2$ ,
- Length term  $\mathcal{L}_g(\phi)$  coefficient:  $\lambda = 1$ ,
- Area term  $\mathcal{A}_g(\phi)$  coefficient:  $\alpha = 5$ ,
- Width of the dirac function:  $\epsilon = 1.5$ ,
- Maximum number of iterations:  $n_{iter,max} = 100$

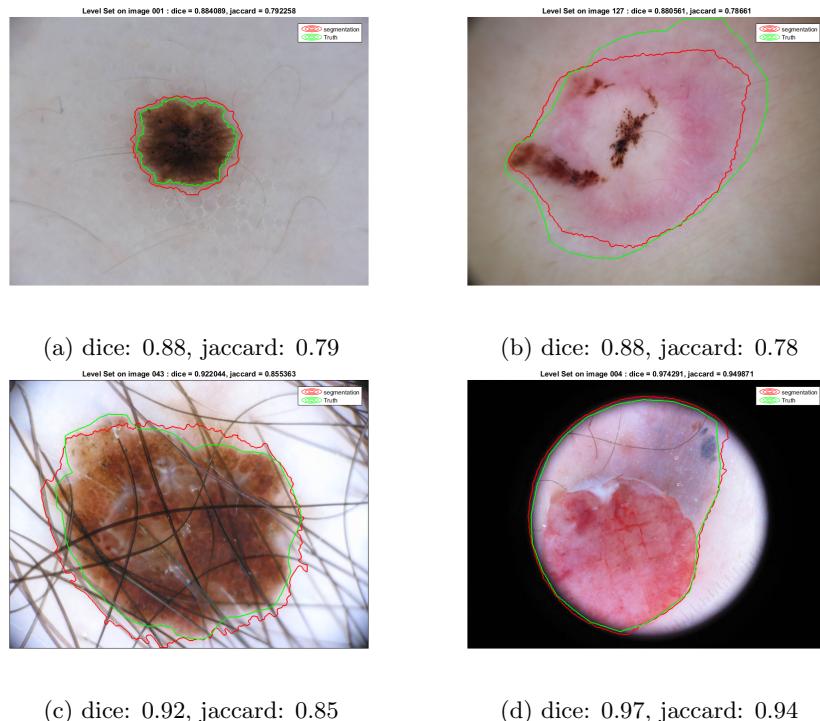


Figure 15: level set

## 5.3 Results

Examples of results are shown in figure 15. We used only the conversion to grayscale and the hair removal preprocessing (no black frame removal). As we can see, the results are good, especially on difficult images such as 15b or 15d, with a dice index of 0.88 and 0.97 respectively, where the previous method were unable to give an acceptable result. However, some "easy" images are oversegmented as illustrated in 15a, but this can probably be enhanced by a two stages evolution as proposed in [9] : after the convergence of the evolution, they set the area coefficient to 0 and compute additional iterations to refine the contour and get closer to the true edges.

Another way to improve my implementation would be to implement a narrow band computation, so that only a small region around the current contour gets updated on each step, to speed up the algorithm, since despite the image cropping, the evolution is still not very fast. This amelioration was however not implemented and the complete dataset evaluation was not made for this method. Please note that even with a fast algorithm, this method would remain the slowest of the three compared here because of the time used for user input.

## 6 Comparison

### 6.1 Performances

As we have seen, the DRLSE provides the best results most of the times, and is especially reliable for difficult images, but its parameters are hard to tune and the segmentation process is slow due to the user input, that must surround the lesion closely enough.

To compare the two other algorithms, we can refer to the figure 16 that overlaps the results from the thresholding method and SRM on the dataset. Although they can differ on certain images, the average performance is the same.

The table 6.1 summarize the advantages and drawbacks of the three methods studied for this project.

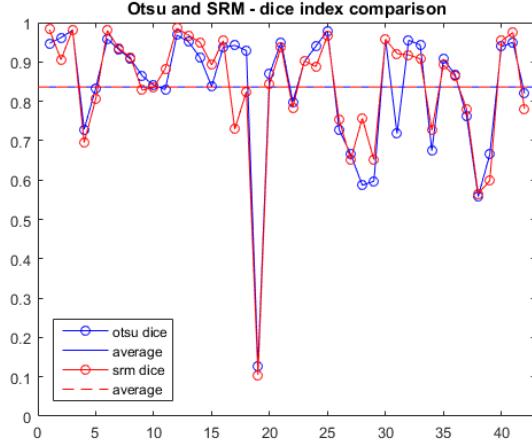


Figure 16: Otsu - SRM comparison : dice index  
The two methods have the same average dice value of 0.84 on the dataset

Method	Thresholding	Region	Level Set
Advantages	simple unsupervised fast no parameter	simple fast	Using prior knowledge precise segmentation No critical failure
Drawbacks	lot of processing	Semi-automated not better than thresholding	Parameters hard to tune slow semi-automated sensitive to initialization

The thresholding have roughly the same performances as the SRM algorithm, and is fully automated. Therefore, it may constitute a better choice for lesion segmentation unless better and more adapted processing can improve the SRM results.

On the other and, the Level Set approach is the best one on difficult images but also the slowest. So it could be interesting to have a choice between the thresholding and Level Set algorithm to provide the best results. A user could review the segmentation results of the thresholding and use Level Set as a semi-automated method on difficult images only.

## 6.2 Choosing the right method

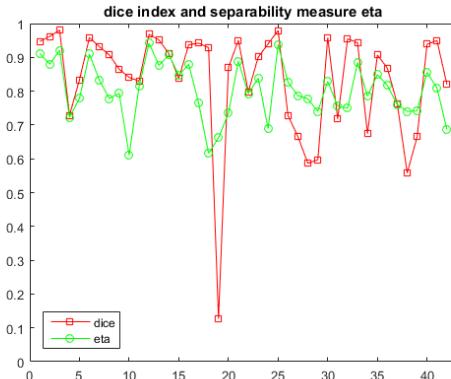


Figure 17:  $\eta$  (green), dice (red)  
Correlation coefficient :  $R(\eta, \text{dice})=0.50$

An interesting development of this comparison would be to have an efficient and easily computable criterion to choose automatically the adapted method given a particular image, so that the operator would not have to review all images. In this section we will show that Otsu's separability measure could make a simple, but not perfect, criterion.

### Separability criterion

To compute the threshold with Otsu's method, we optimize a separability measure  $\eta$ . We investigated whether there were a correlation between this value and

the evaluation metrics, so that it could be used to evaluate the performance of the thresholding algorithm. The results are shown in figure 17 and it appears that  $\eta$  is correlated to the evaluation metrics with a correlation coefficient of 0.5.

To test this criterion, we determined on a training set a threshold for eta that would ensure a good performance ( $dice > 0.7$ ) for the thresholding method. That could help us to automatically detect good from bad performances on real data and for instance switch to a semi-automatic method in that case. We also wanted to control how many images would have to be rejected. Given the algorithm performances described in section 3.3, we should only drop the 7/42 (16.7%) of the dataset where the dice index was lower than 0.7.

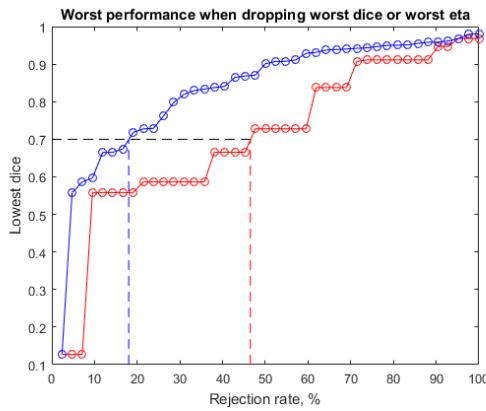


Figure 18: Evolution of the performance when dropping bad results  
 Red curve : dropping images with the lowest separability metric  $\eta$   
 Blue curve : dropping images with the actual worst dice value  
 Dotted lines : percentage of image rejected to ensure  $dice > 0.7$

The figure 18 shows the evolution of the lowest performance while we iteratively drop the image with the smallest  $\eta$  value, ie what we believe to be the worst performance. We compared that to the evolution of the worst performance while dropping the image with the smallest dice index, ie the actual worst performance. To ensure a dice index higher than 0.7 on this training set, we would have to reject all images with  $\eta < 0.8$ , which represent about 47% of the set. This rejection rate is high, but if we lower the  $\eta$  threshold to 0.7, decreasing the rejection rate to a more acceptable 11%, the lowest dice index accepted as "good" comes down to 0.55.

Therefore, the selected threshold for  $\eta$  is 0.8.

To test how reliable this technique is, we evaluated this 0.8 threshold on a test set of 50 images. The figure 19 shows that the criterion ensured indeed a dice index above 0.7 (only one outlier at 0.68), at the cost of 42% (21/50) rejection rate.

## Ground Truth quality

The ground truth used was not perfect: it is hard to produce segmentation gold standard, especially in lesion segmentation because of fuzzy borders, color and

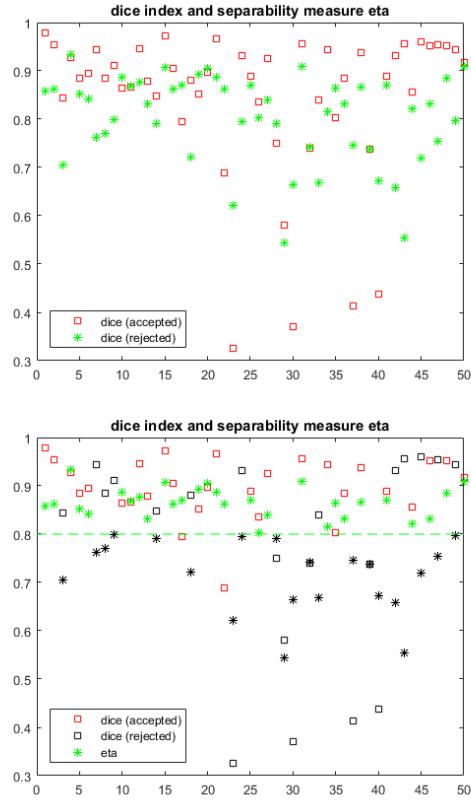


Figure 19: Test of the criterion - dice : squares,  $\eta$  : stars, rejected : black

shape variability and other defaults. In this dataset from ISIC, two different methods of manual segmentation were used and our algorithms were not easily able to perform well for both.

## Conclusion

We have shown in this project that the adaptative thresholding can give good result when associated with pre- and post-processing. The SRM method that we implemented would need some improvement to justify the need of a user interaction in the process by a real increase of performance, compared to the thresholding. Level Set is the most reliable method for difficult images, but require tuning many parameters and we have probably not chosen the optimal values, but the drawback is a high runtime and a high cost in terms of user interaction. Therefore it should be only used on difficult images.

We propose a criterion to automatically detect such images using the separability measure used in Otsu's optimal threshold selection. This criterion has a high sensitivity but a low specificity, so it comes with a high cost of wrongly rejected images. However, a user would only need a short time to flag a rejected image as "acceptable".

## References

- [1] M. Emre Celebi, Hitoshi Iyatomi, Gerald Schaefer, and William V. Stoecker. Lesion border detection in dermoscopy images. *Computerized Medical Imaging and Graphics*, 33(2):148–153, March 2009.
- [2] M. Emre Celebi, Hassan Kingravi, Hitoshi Iyatomi, Y Alp Aslandogan, William Stoecker, Randy Moss, Joseph M Malters, James Grichnik, Ashfaq A Marghoob, Harold S Rabinovitz, and Scott W Menzies. Border detection in dermoscopy images using statistical region merging. *Skin research and technology : official journal of International Society for Bioengineering and the Skin (ISBS) [and] International Society for Digital Imaging of Skin (ISDIS) [and] International Society for Skin Imaging (ISSI)*, 14:347–53, 2008.
- [3] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Lipopyris, Nabin Mishra, Harald Kittler, and Allan Halpern. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). *arXiv:1710.05006 [cs]*, October 2017. arXiv: 1710.05006.
- [4] Jianbo Gao, Jun Zhang, M. G. Fleming, I. Pollak, and A. B. Cognetta. Segmentation of dermatoscopic images by stabilized inverse diffusion equations. In *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269)*, pages 823–827 vol.3, October 1998.
- [5] R. Garnavi, M. Aldeen, M.E. Celebi, A.Bhuiyan, C. Dolianitis, and G. Varigos. Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels. *International Journal of Medicine and Medical Sciences*, 1(2):126–34, 2010.
- [6] David Delgado Gómez, Constantine Butakoff, Bjarne Kjaer Ersbøll, and William Stoecker. Independent histogram pursuit for segmentation of skin lesions. *IEEE transactions on bio-medical engineering*, 55(1):157–161, January 2008.
- [7] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder. Diagnostic accuracy of dermoscopy. *The Lancet Oncology*, 3(3):159–165, March 2002.
- [8] Tim Lee, Vincent Ng, Richard Gallagher, Andrew Coldman, and David McLean. Dullrazor®: A software approach to hair removal from images. *Computers in Biology and Medicine*, 27(6):533–543, November 1997.
- [9] Chunming Li, Chenyang Xu, Changfeng Gui, and Martin D Fox. Distance regularized level set evolution and its application to image segmentation. *IEEE transactions on image processing*, 19(12):3243–3254, 2010.
- [10] Teresa Mendonça, André Marçal, Angela Vieira, Jacinto Nascimento, Margarida Silveira, Jorge Marques, and Jorge Rozeira. Comparison of Segmentation Methods for Automatic Diagnosis of Dermoscopy Images. *Conference proceedings : ... Annual International Conference of the IEEE Engi-*

neering in Medicine and Biology Society. *IEEE Engineering in Medicine and Biology Society. Conference*, 2007:6573–6, February 2007.

- [11] Richard Nock and Frank Nielsen. Statistical Region Merging. *IEEE transactions on pattern analysis and machine intelligence*, 26:1452–8, December 2004.
- [12] N. Otsu. A threshold selection method from gray-level histograms. *Man and Cybernetics, IEEE Transactions on Systems*, 9:62–66, 1979.
- [13] P. Schmid. Lesion detection in dermatoscopic images using anisotropic diffusion and morphological flooding. In *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, volume 3, pages 449–453 vol.3, 1999.
- [14] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IEEE Transactions on Medical Imaging*, 36(4):994–1004, April 2017.