Università degli Studi di Cagliari

FACOLTÀ DI SCIENZE
Corso di Laurea Magistrale in Informatica

White blood cells segmentation using Vector Field Convolution

Relatore
Prof. Cecilia Di Ruberto

Candidato
Simone Porcu

Anno Accademico 2016-2017

To my family

Contents

Introduction

Blood is a body fluid deliver. It's contains an transport many of the nutrients substances that the man and the other animals use to live. That we call blood is principally a fluid divided in two elements: blood cells and blood plasma. Normally an individual has around 5 Litre of blood. The plasma blood constitutes the 55% of the total fluid. it is mostly water(92% by volume) and contains proteins, glucose, mineral ions, hormones and blood cells themselves.[3] Mainly the cells are red blood cells and white blood cells(WBCs). In this dissertation we going to focus on WBC especially we will study the shape of these last. White blood cells, also called leukocytes , are the cells with the task of controlling the body against both infectious disease and foreign invaders. All leukocytes have a nuclei that distinguishes them by other blood cells, in particular red blood cells and platelets. The generic term leukocytes includes very different cells population: neutrophil, granulocytes, basophilic granulocytes and eosinophilic granulocytes. This set of three categories is defined as polymorphonucleated granulocytes. The other set that includes monocytes and lymphocytes is defined agranulocytes mononuclear. In a nutshell leukocytes are dived in these two sets by the nuclei shape.1

White blood cells segmentation

Segment an image means divide an image in regions of interest. It's used to obtain a more compact image used to extract objects or to analyse an image. In this case the main feature is to find edges and white blood cells nuclei. At a first look seems a banal problem, because the we think that every single cell is strongly separated by the others, but obviously it is the best case that we can find. Commonly the microscope photos that we analyse contains noise and in particular the leukocytes overlaps both others leukocytes and red blood cells. For these reasons segment leukocytes is still an unresolved problem. As explain above there are 2 class of leukocytes that are dissimilar by the nuclei shape. this is an high problem because if the solution to find all the white blood cells was based on the search of circular shapes, it's trivial that it will be impossible to recognize a granulocytic from a monocyte.

There are a lot of heuristics and approaches that try to divide and classify

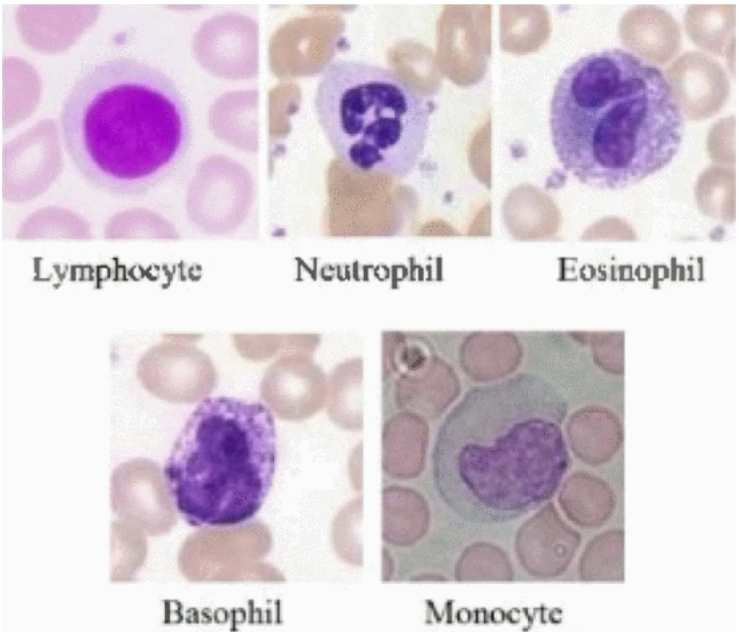


Figure 1: Example of the different kind of leukocytes[2]

white blood cells. This dissertation proposes a new approach of pure segmentation using the Vector Field Convolution, in particular tries to find a division between the overlaps between the cells. We choose this kind of field because the common practice to extract the features by the images utilizes thresholds, but what happens if the image has a low definition and all the cells are in overlap with their? Using that field to describe the image is possible to transcend by the shape of the features and focus themselves on the points that have a non-uniform virtual field. This technique then considers only the points that describe the edges of the white blood cells. After the image elaboration the result is an image that contains only leukocytes and when it is necessary dividing every kind of cells. With this result we can label every cell without human work.

Part I

Background

Chapter 1

An overview about vector fields

Active contours, also called snakes, are curves that move inside the image following the energy of the field. There are two kinds of forces, one internal and anther external. Combining these two it's possible to create a curve that follows constraints gives by the forces. The internal and external forces are defined so that the snake will conform to an object boundary or other desired features within an image. Snakes are widely used in many applications, including edge detection, shape modelling and segmentation. There are two general types of active contour models in the literature today: parametric active contours and geometric active contours. Typically, the curves are drawn toward the edges by potential forces, which are defined to be the negative gradient of a potential function. Additional forces, such as pressure forces, together with the potential forces comprise the external forces. There are also internal forces designed to hold the curve together and to keep it from bending too much. There are two levels difficulties with active contour algorithms. First, the initial contour must be close to the true boundary or else it will likely converge to the wrong result. The second problem is that active contours have difficulties progressing into concave boundary regions. Although many methods such as multi resolution methods, pressure forces, distance potential forces, control points, and using solenoidal external fields have been proposed they either solve one problem or solve both but creating new difficulties. For example, multi resolution methods have addressed the issue of initialization, but specifying how the snake should move across different resolutions remains problematic. Another example is that of pressure forces, which can push an active contour into boundary concavities, but cannot be too strong or "weak" edges will be overwhelmed. But how works a snake if the objects to segment are overlapped? Snakes are able to find all the external edges of the object but in this case the edge can be consider an internal part of the object. With the active contours is impossible segment the overlapped cells because the snake cannot enter inside the cell region. For these reason we have used our virtual

field following another lecture key.

Part II

A parallel way of using VFC without using active contours

Chapter 2

The implementation

2.1 Vector field convolution

Convolving a vector field with the edge of the map derived from the image you get an external force, the VFC. Active contours using the VFC external force are called VFC snakes. Like the GVF snakes instead of being Formulated using the standard energy minimization framework, VFC snakes are constructed from a state of equilibrium between the forces. The VFC snakes besides having a wide capture range and the ability to capture the concavities, are better resistant to noise image, have the ability to adapt the force field and reduce drastically the computational cost. Before to explain the VFC is right explain the vector field kernel

$$k(x, y) = m(x, y)n(x, y) \tag{2.1}$$

where n is the unit vector that points to the origin of the kernel

$$n(x, y) = [\frac{-x}{r}, \frac{-y}{r}] \tag{2.2}$$

and m is the magnitude of the vector . The authors of the VFC implemented two kind of magnitude. If we consider the origin as the point of interest, this vector field kernel has the desirable property that a free particle placed in the field is able to move to the point of interest. The external force that work in the VFC is defined in this way:

$$f_{vfc}(x, y) = u_{vfc}(x, y), v_{vfc}(x, y) \tag{2.3}$$

Since the map of the edge is non-negative and is wider near the edges of the image, the edges act to a greater extent on the VFC than homogeneous regions. Therefore, the free particles of homogeneous regions will be attracted to the edges. If we present the vector field kernel using a complex-valued range, the VFC is just the filtering result of the edge map, which does not depend on the origin of the kernel. The VFC field highly depends on the magnitude of the vector field kernel . Field VFC has the magnitude directly proportional

to the vector field kernel (x, y). Knowing that the figure of interest has less influence on the particles away from it, the magnitude must be expressed as a positive function decreasing with respect to the distance of the origin. Below are propose two types of magnitude functions, given as

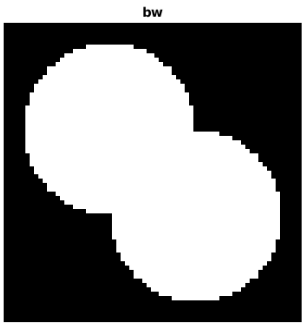
$$m_1(x, y) = (r + \epsilon)^{-\gamma} \tag{2.4}$$

$$m_2(x, y) = exp(-r^2\,\zeta^2) \tag{2.5}$$

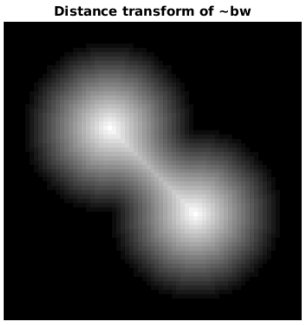
where γ and ζ are positive parameters to control the decrease, ϵ is a small positive constant to prevent division by zero at the origin. $m_1(x, y)$ is inspired by Newton's law of universal gravitation in physics. Furthermore, the pixels in the edge map can be considered as objects of mass proportional to the strength of the edges and the field VFC would be the gravitational field generated by all objects. The influence of the figure of interest increases as γ decreases. In practice γ usually ranges from 1.5 to 3 for most images. $m_2(x, y)$ is a Gaussian shape function, where ζ can be viewed as the standard deviation. The influence of the figure of interest increases as ζ increases. In general, the influence of the figure of interest should be increased (decrease or increase) as the signal-to-noise ratio is decreased.[1]

2.2 New kind approach to segment leukocytes

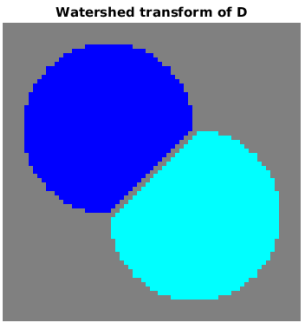
As we can see in literature, the main approach used to resolve the overlap problem is using the watershed transform. Here2.1a there is an example of the watershed transform. Takes in input the image of the two overlapped circles, we calculate the distance transform, or in other words we calculate the Euclidean distance transform of the binary image BW. For each pixel in BW, the distance transform assigns a number that is the distance between that pixel and the nearest nonzero pixel of BW2.1b.Giving the distance transform result to the watershed algorithm we obtain a division between circles because the watershed transform finds "catchment basins" or "watershed ridge lines" in an image by treating it as a surface where light pixels represent high elevations and dark pixels represent low elevations.2.1c This is an ideal case to analize. But when we work with the overlapping between cells the result of the division by Watershed transform is not optimal like the example 2.1. Probably the cause derived from the low definition of the figures and especially the shape of the cells. Here there is an example of what happened when we try to divide 3 cells in overlap2.2. As is possible to see, this method produce a no-realistic separation of the cells. For this reason we study a different method that in automatic way produce a realistic division of the cells. The algorithm uses principally the output of the VFC field, the image relative to the External energy of the image to describe the edges, the median filter and the skeleton method.



(a)

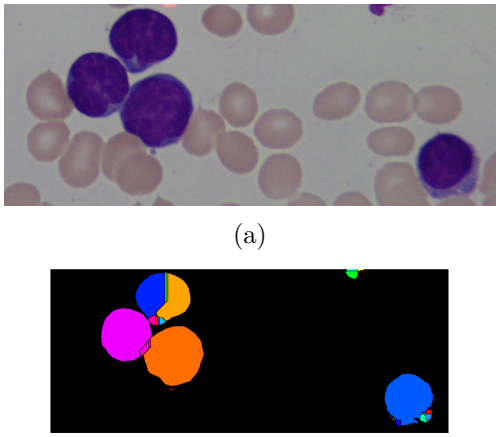


(b)



(c)

Figure 2.1: (a) Example of two circle in overlap, (b) Distance transform, (c) Watershed result



(a)

(b)

Figure 2.2: (a) Original leukocytes image, (b) Watershed transform applied to three cells in overlap

2.3 The VFC result

The VFC uses the two components of the external force $u_{vfc}(x,y), v_{vfc}(x,y)$ to describe the field of the image and it's magnitude. Our purpose is find and image using these two components that describes all the leukocytes edges taking an accurate look on the edges in overlap. The first step then is extract the right component and the left component.

$$u_{vfc} = ExtF(x)/\sqrt{ExtF(x)^2 + ExtF(y)^2} \tag{2.6}$$

$$u_{vfc} = ExtF(y)/\sqrt{ExtF(x)^2 + ExtF(y)^2} \tag{2.7}$$

where ExtF is the External force of the field. Now we have only an intensity image, but to understand how the field moves in the space we have to transform these two component u and v in grades. It a mandatory do this step because we want to understand the direction of every pixel in the figure. Is possible convert the two components in degrees using the $atan2d$ function 2.3. In order to delete all the uniform part of the figure and put in exalt the edges we use a mediand filter using the function $ordfilt2$ searching the 18th element of the $5 * 5$ mask 2.4.

2.4 External Energy

As is possible to see in the result 2.3, there are a lot of points that are artefacts created by the field. For these reason we used the *bwdist* function to assign

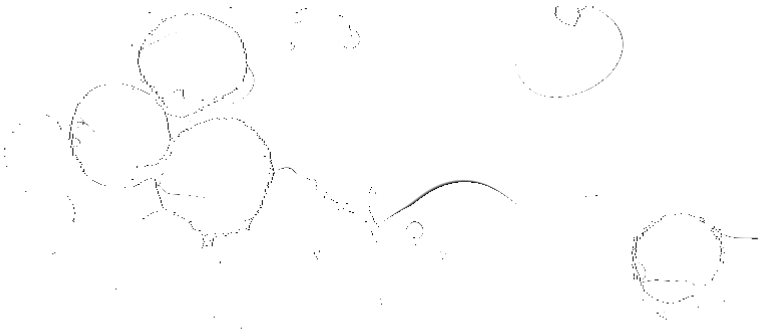


Figure 2.3: degrees image

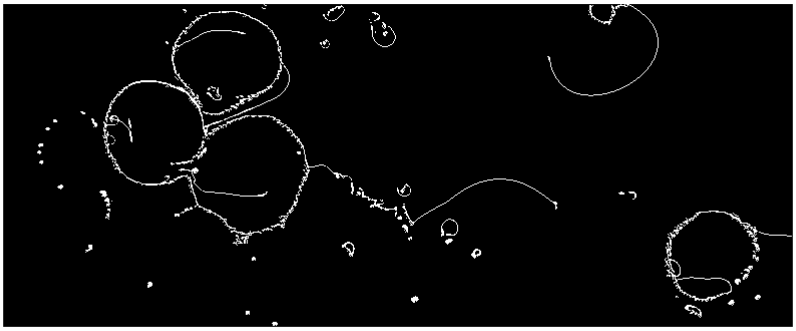


Figure 2.4: Median filter on degrees image

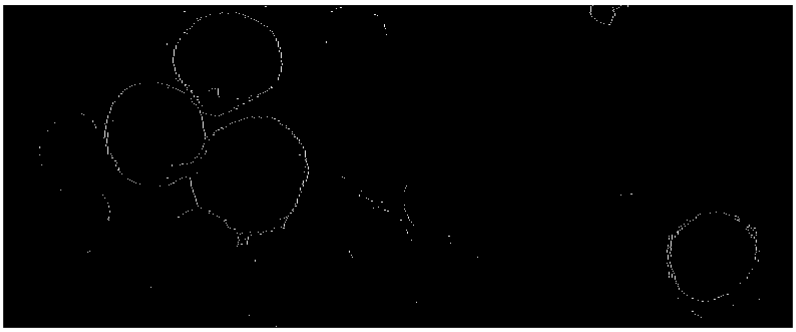


Figure 2.5: bwdist applied on degrees image

a number that it is the distance between each pixel and the nearest no-zero pixel of the image. This trick is very useful because reduce the entropy of the image, focusing only on the shape of the leukocytes 2.5. But we have ever the same problem. in the image there are trace of the red blood cells, then we had to find a method to isolate only the leukocytes. We started using the external energy of the image. For an image $I(x, y)$ all the lines, edges and terminal points the general formulation of the Energy of the image is

$$E_{image} = w_{line}E_{line} + w_{edge}E_{edge} + w_{term}E_{term} \tag{2.8}$$

where $w_{line}, w_{edge}, w_{term}$ are weights of the features.

2.4.1 Line functional

The line functional or in other terms the intensity of the image is in a nutshell the attracted value of the dark lines to the light line. It's possible choose this attraction putting a positive or negative sign before the force that this attraction has to be.

$$E_{line} = filter(I(x, y)) \tag{2.9}$$

2.4.2 Edge functional

The edge functional bases it's work on the image gradient.

$$E_{edge} = -\left|\nabla I(x, y)\right|^2 \tag{2.10}$$

It's very useful because when we try to analyse the feature of the image, we work with maxims and minims. with this formula we can avoid the local

minima that are not object of interest. The energy functional using scale space continuation is

$$E_{edge} = -\left|G_\sigma * \nabla^2 I\right|^2 \tag{2.11}$$

where G_σ is a Gaussian with standard deviation σ.

2.4.3 Termination functional

The curvature of the lines in a image is utilized to detect corners and terminations. Put

$$C(x,y) = G_\sigma * I(x,y) \tag{2.12}$$

with a gradient angle

$$\theta = \arctan\left(\frac{C_y}{C_x}\right), \tag{2.13}$$

unit vectors that move along the gradient direction

$$\mathbf{n} = (\cos\theta, \sin\theta) \tag{2.14}$$

and unit vectors perpendicular to the gradient direction

$$\mathbf{n}_\perp = (-\sin\theta, \cos\theta). \tag{2.15}$$

With these 4 equations we can describe the termination functional of energy as follow

$$E_{term} = \frac{\partial\theta}{\partial n_\perp} = \frac{\partial^2 C/\partial^2 n_\perp}{\partial C/\partial n} = \frac{C_{yy}C_x^2 - 2C_{xy}C_xC_y + C_{xx}C_y^2}{(C_x^2 + C_y^2)^{3/2}} \tag{2.16}$$

2.4.4 External energy result

Now we have only to specify the value of each parameter explained above. To obtain the desired outcome we did some empirical experiments, obtaining the best result with $Wedge = 8, Wline = -8 and Wterm = 0$. this result 2.6 permit us to extract only the leukocytes part of the image using some analysis image exploit 2.7.

In order to delete all the uniform part of the figure and put in exalt the edges we use a mediand filter using the function $ordfilt2$ searching the 18th element of the $5 * 5$ mask 2.8.

Figure 2.6: External energy result

Figure 2.7: image of leukocytes without red blood cells

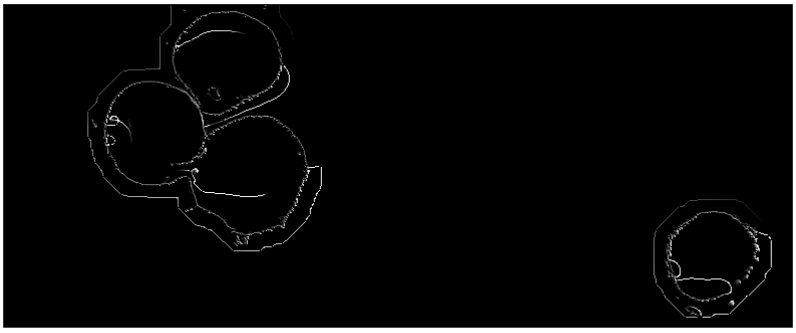


Figure 2.8: edges and region of leukocytes

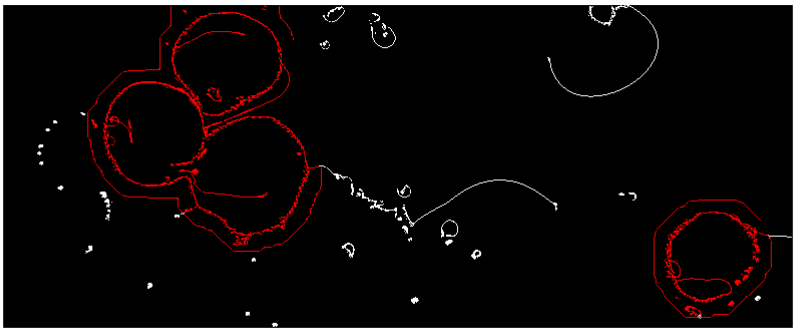


Figure 2.9: overlay of figures 2.4 and 2.8

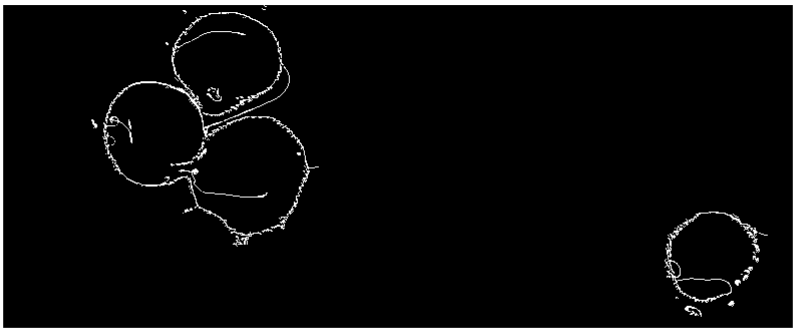


Figure 2.10: only leukocytes edges

2.5 Combination of two results: the division method

The two result that we obtained seems in no-correlation, but the skill of this segmentation live in this passage. Using the overlay function we search all the points in overlay between the images 2.4 and 2.8, using the red color to isolate the leukocytes region 2.9. We choose to use the red color, because if we isolating only the red component of the image we can obtain an image that contains only the leukocytes regions. The result of this task is visible in the image 2.10

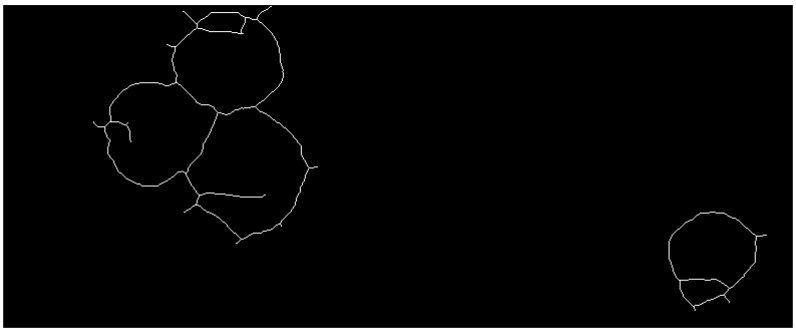


Figure 2.11: skeleton of leukocytes with irregular branches

2.6 Segmentation with the skeleton function

Starting from the image 2.10 we have to connect every single white point to the nearest. To do this task we use the function *imdilate* to dilate all the white dots with a diamond structural element with the size of 6 pixels. after this step we apply the closing of the opening with two disk respectively of the size 3 and 4 pixels. Now we can apply the skeleton function or the thinning of the edges. To do this passage we use the Matlab function *bwmorph* that it's not the best function to do this but is the faster one. We try another skeleton external function written by N. Howe that is very interesting because has a precision to compute the skeleton of the image that is very impressive, but because the Pc latency we can't use the last one function. After the skeleton application we obtain an image that contains some spurious branches 2.11. To solve this problem is possible to use the Matlab function *bwmorph* to prune the spurious branches, but in this case the function doesn't work very well. For this reason we implement a code to resolve the problem. Our code is like a parser that for each point it's see if it is a part of a close circle or not. If it's not a part of a loop the code deletes this point. In a nutshell we save only the point that stay in a 'road' with the starting point and the end point coincident. Below you can read a snippet of pruning code.

```
B = branchpoints;
E = endpoints;
[y,x] = Image;
Dmask = false(size(skel));
        for k = 1:numel(x)
        D = bwdistgeodesic(skel,x(k),y(k));
        distanceToBranchPt = min(D(B));
        Dmask(D < distanceToBranchPt) =true;
        end
skelD = skel - Dmask;
```

The result that we obtain from the skeleton application is a summary work, indeed we obtain an over-segmentation as is viewable from the image 2.12. This over-segmentation isn't good for the correct visualization of leukocytes,

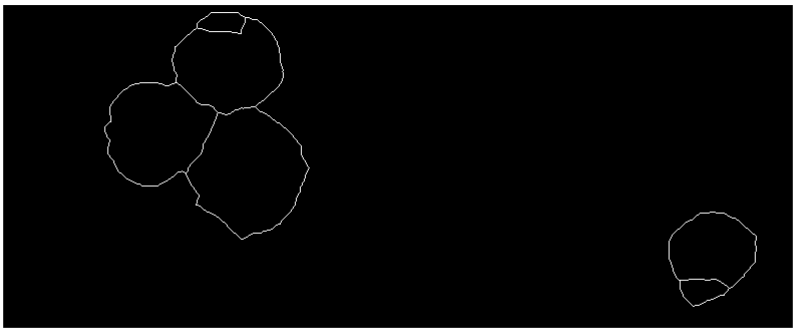


Figure 2.12: skeleton of leukocytes with no spurious branches

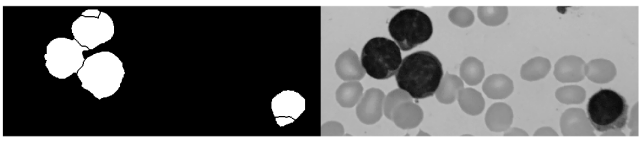
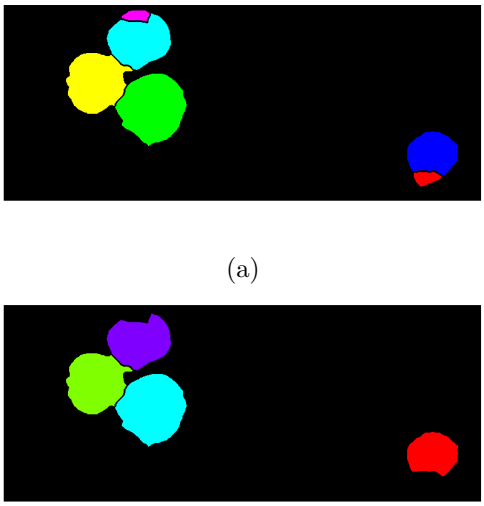


Figure 2.13: final leukocytes segmentation

but gives us a starting point to improve the solution and to obtain a better result. Then we try to combine the various result to obtain a segmentation that was similar to the original image. First of all we close all the holes that are inside the image, to obtain a sort of black and white mask. Another fundamental step is sum the skeleton image with the mask. Doing this step we can separate the foreground by the background. Now we can use for the last time the map of edges that we used to calculate the VFC field. We use it summarizing it with the image of leukocytes in foreground and passing this image sum as input to the function $bwareafilt$. We do this because this function extracts all connected components (objects) from a binary image where the area is in range, producing the segmented image of the leukocytes 2.13.

2.7 Cells counting

To understand if all that we did had a sense, we have to do a counting of the cells. Because the images have a low definition we can found an over segmentation inside the cells, but it's easy to overtake this problem without consider the little regions that are inside the image. Then if we delete all the regions that are less than an upperbound we can have the exact number of



(a)

(b)

Figure 2.14: (a) All the regions, (b) Only big regions

leukocytes inside the image. We do this using the following snippet code

```
CC = bwconncomp(BW2,8);
numPixels = cellfun(@numel,CC.PixelIdxList);
[~,idx] = min(numPixels);

while min(numPixels)<2000
    BW2(CC.PixelIdxList{idx}) = 0;
    CC = bwconncomp(BW2,8);
    numPixels = cellfun(@numel,CC.PixelIdxList);
    [~,idx] = min(numPixels);
end
[labeledImage, numberOfObject] = bwlabel(BW2);
```

The resulting image of this consider only the the regions that are the same size of the nucleus or bigger.

Appendices

Appendix A

Appendix chapter

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas arcu ipsum, posuere sit amet condimentum eu, accumsan id tortor. Nam libero ex, eleifend nec felis vel, dignissim rutrum nisl. Donec et lorem non urna rutrum viverra. Vivamus mollis quam lacus, consectetur finibus felis iaculis volutpat. Praesent egestas, enim vel ultricies dictum, neque felis eleifend velit, ut pulvinar leo lorem at orci. Suspendisse potenti. Donec quis ligula diam. Fusce urna justo, consequat et justo non, pellentesque fermentum nisl. Mauris vitae purus id purus luctus tincidunt. Integer congue velit ut dui feugiat lacinia. Quisque rutrum sapien at orci sodales lacinia. Donec congue placerat nulla, at euismod urna dignissim id. Aliquam feugiat euismod aliquam. Nulla facilisi.

A.1 Section

Aenean sit amet pretium purus. Suspendisse vel arcu nec ex tristique euismod vitae vel lacus. Nunc tristique ante in congue mattis. Donec nec augue sed est sagittis dapibus et vel nisl. Curabitur facilisis posuere rhoncus. Pellentesque vitae malesuada ipsum. Aenean blandit sem non gravida auctor. Ut sed eleifend ex. Aenean pellentesque ornare auctor. Aliquam maximus sem neque, at pharetra tortor iaculis vitae. Nulla aliquet diam quis lacus semper, eget suscipit tortor vestibulum. Ut non dignissim sem. Curabitur tincidunt, ex et luctus rhoncus, elit augue rutrum erat, at pellentesque enim lectus eu nulla. Nam varius leo mollis nisl tincidunt, eu vestibulum ante maximus. In condimentum dictum pellentesque.

Donec feugiat magna vel quam rhoncus, ut sodales purus lobortis. Nam vestibulum varius arcu, ut tristique est lacinia et. Sed efficitur dictum elementum. Aliquam a ultricies nisl. Vestibulum hendrerit ante ligula, et finibus tortor volutpat non. Donec tristique sapien vel egestas laoreet. Proin lobortis, ipsum quis hendrerit interdum, tellus leo efficitur leo, non mollis libero turpis vitae nunc. Duis sed mollis augue. Integer in purus non lectus interdum hendrerit. Curabitur venenatis velit orci, nec bibendum velit convallis et. Quisque sodales feugiat luctus. In sed nunc tempus, egestas ligula nec, tristique augue.

Fusce tempor urna mi, vitae maximus lectus sodales ut.

List of Figures

List of Tables

List of listings

Bibliography

[1] Scott T. Bing Li. "Active Contour External Force Using Vector Field Convolution for Image Segmentation". In: *Acton* (2007) (cit. on p. 10).

[2] Nikhil Vadera Shubham Manik Lalit Mohan Saini. "Counting and classification of white blood cell using Artificial Neural Network (ANN)". In: *IEEE* (2017) (cit. on p. 2).

[3] Wikipedia. *Blood*. 2007. URL: `https://en.wikipedia.org/wiki/Blood` (cit. on p. 1).