Automatic segmentation of arteries in multi-stain histology images

L. Leal Taixé[1], A. U. Coskun[2], B. Rosenhahn[1] and D. H. Brooks[2]

[1] Leibniz Universität Hannover, Hannover, Germany
[2] Northeastern University, Boston, USA

Abstract— **Atherosclerosis is a very common disease that affects millions of people around the world. Currently most of the studies conducted on this disease use Ultrasound Imaging (IVUS) to observe plaque formation, but these images cannot provide any detailed information of the specific morphological features of the plaque. Microscopic imaging using a variety of stains can provide much more information although, in order to obtain proper results, millions of images must be analyzed. We provide an automatic way to find the Region of Interest (ROI) of these images, where the atherosclerotic plaque is formed. Once the image is well-segmented the amount of fat and other measurements of interest can also be determined automatically. A proper study of these images provides valuable information that can be key to understanding the evolution of atherosclerosis.**

Keywords— **automatic segmentation, ROI, statistical color model, snakes, splines**

I. Introduction

Coronary Heart Disease (CHD), also called Atherosclerotic Heart Disease or Coronary Artery Disease (CAD) is a serious disease caused by the formation of plaque in the arteries. There are two common approaches for plaque determination; (a) Invasive, in-vivo imaging techniques such as Ultrasound Imaging (IVUS) [1], Optical Coherence Tomography (OCT) [2] or Magnetic Resonance Imaging (MRI), which suffer from noise induced by dynamic motions and low resolution. (b) Microscopic histology imaging of samples ex-vivo, which has a much higher resolution and allows for a detailed study of the morphology of the plaques. All the above techniques need an appropriate automatic segmentation tool to analyze large image databases in order to improve the understanding of the disease. Currently, such images are usually analyzed manually, e.g. using Photoshop®, which is a time consuming and error-prone process. The present study focuses on the automatic segmentation of stained microscopic images to analyze coronary plaque characteristics. The contributions can be summarized as follows:

1. An automatic inner and outer boundary detector is introduced to capture the area of the artery within the image.
2. We propose to create a statistical color model derived from training data and we compensate for color differences created during the staining process.

3. Our approach automatically detects starting points for contour evolution.
4. We design a Graphical User Interface (GUI) that dramatically reduces the workload of the user.

Figure 1 shows the basic parts of an artery, along with a brief description of each. We are interested in capturing the
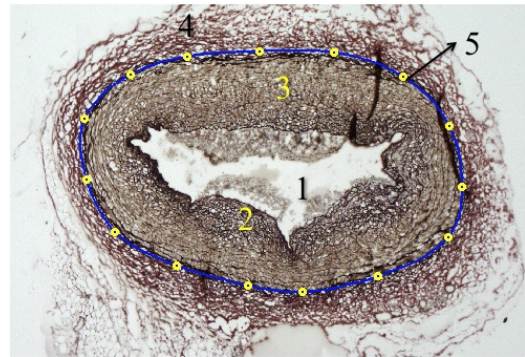

Fig. 1: 1.*Lumen*: opening inside the artery where the blood flows. 2.*Intima*: first layer, made of a single layer of endothelial cells. This is where the lipids are deposited to form an atherosclerotic plaque. 3.*Media*: second layer, made of smooth muscle cells and elastic tissue. 4.*Adventitia*: third and outermost layer, mainly composed of collagen which gives stability to the blood vessel. 5.*External Elastic Lamina (EEL)*: wall between the media and the adventitia.

plaque as well as the area of the artery (intima and media). As the atherosclerotic plaque is formed in the intima, the target ROI can be found between the EEL (outer boundary of the ROI) and the contour of the lumen (inner boundary of the ROI).

The basic structure of our algorithm is shown in Figure 2. In a timely fashion, we can divide the algorithm in:

- Creation of a color model from training data. This is computed just once and is marked in blue.
- Finding the EEL using snakes with the information of the likelihood image and the inner boundary.
- Feedback of the GUI which affects the inner boundary determination and the cleaning of the likelihood image. Marked in red.
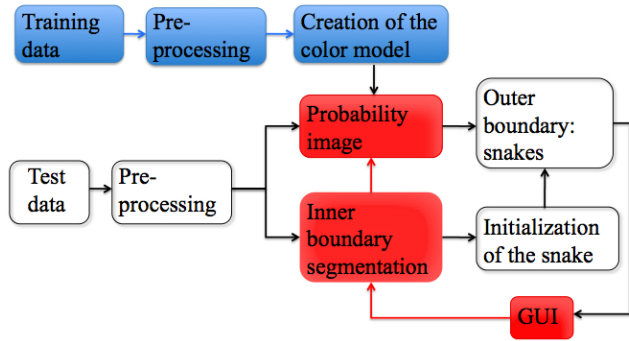
Fig. 2: Diagram of the algorithm

The remainder of the paper is organized as follows: Section 2 presents the database, basic pre-processing and inner boundary segmentation. Section 3 presents the creation of the statistical color model and Section 4 the automatic detection of the EEL. Finally, the results and conclusions are presented.

II. PREPARATION OF THE IMAGES

A. An overview of the database

For this study we use 42 histological images from 4 different coronary artery types: the Right Coronary Artery (RCA), the Left Circumflex Artery (LCX), the Left Anterior Descending Artery (LAD) and the Obtuse Marginal Artery (OM) within a total of 33 arteries from 11 different pigs.

We focus on images created with the stains:

- Verhoeff: darkens the structural elements of the artery (IEL, EEL). We use these images to find the initial ROI.
- Oil Red O: shows in bright red the fat in the artery. We use these images to compute the total amount of plaque.

We then use color features as in [3, 4] to create a model from the training data.

Our study faces 3 main problems with the images (see Figure 3). Each one will be dealt with a specific method:

1. Color differences due to the staining process.
2. External tissue or stain spots that result in undesired white artifacts in the likelihood image.
3. Various shapes and sizes of the arteries.

B. Pre-processing

In order to compensate for color variability (problem 1), we decided to use histogram equalization. The obvious approach of using three different color models (pink, brown and purple) proved to give less consistent results in the end.

C. Inner boundary: region growing

To segment the inner boundary of the ROI, we use the region growing technique [5]. We choose the seed manually and apply the algorithm on a grayscale image. It returns a closed region with all the pixels with values $\pm T$ with respect to our seed pixel intensity value, where T is typically 40.

III. STATISTICAL COLOR MODEL

The outer boundary of our ROI can be clearly identified by a dark color produced by the Verhoeff stain. Therefore, we model this color information and create a likelihood image that indicates where in the image it is likely to have an EEL.

A. Model training and creation

To generate the statistical color model, we manually select $N_t = 15$ boundary points on all the training images. We include in the model all the pixels in a window w of 11x11 centered in the selected boundary point.

A normalized histogram with M bins is created for each of the color components R, G and B of the pixels.

B. Likelihood image

In order to convert the histogram information to the pixels in the image space, we create what we call the likelihood image. As shown in Equation (1), we use a naive Bayes classifier to compute the probability of each pixel of the image.

$$P(x,y) = P(x_R, y_R | M_R) * P(x_G, y_G | M_G) * P(x_B, y_B | M_B) \quad (1)$$

According to the model, the likelihood image should be white in the boundary pixels and black elsewhere.

In order to solve problem 2 (undesired white pixels in the likelihood image), we clean the image using morphological operators as in [6, 7, 8]. We apply image reconstruction [9] as expressed in Equation (2), using an eroded version of our image as the marker. One example is shown in Figure 4.

$$\delta(...\delta(\delta(Y) \wedge X)... \wedge X) = (\delta_1^X(Y))^\infty = \delta_\infty^X(Y) \quad (2)$$
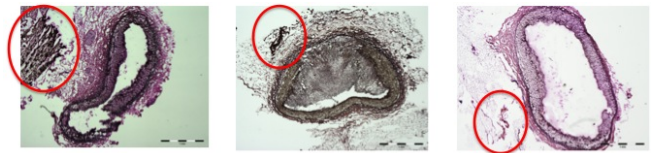

Fig. 3: Problem 1,3: color and shape differences – Problem 2: marked in red

We decide for a dynamic size of the structuring element (E_{size}) for the erosion, as shown in Equation (3).

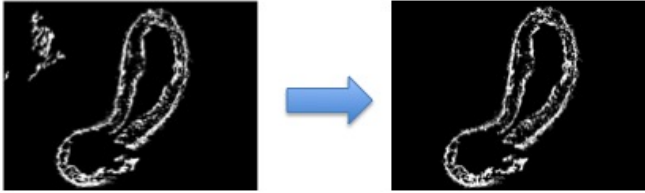$$E_{size} = 50 - 5 * i \qquad i = 0 \ldots 10 \qquad (3)$$

Fig. 4: Image reconstruction effect on likelihood image

IV. OUTER BOUNDARY: SNAKES AND SPLINES

In order to find the outer boundary of the arteries, we need an adaptive contour that is able to "ignore" the white artifacts that might still remain in some images but that is also able to "imagine" where the boundary is when the contour is not closed. We use snakes or active contours expressed using splines [10]. The snakes allow us to "drive" our contour to the highest energy point. We use splines to exploit the fact that arteries have rounded shapes, so our contour must have a polynomial behavior where no sharp edges are possible.

A. Automatic initialization

One of the most important steps to see our segmentation algorithm succeed, is finding a good initialization for the snake. The initialization depends on the shape and size of the artery (problem 3), therefore we need to use known contour information. Our approach is to use the inner boundary found previously using region growing. As shown in Figure 5, we
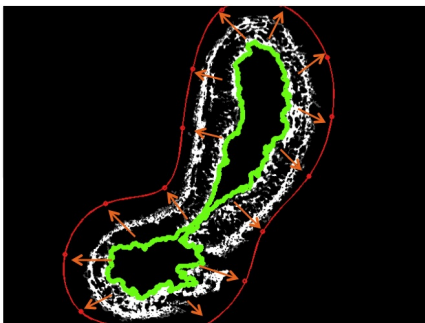

Fig. 5: Finding the snake starting points

start from the inner boundary (green) and dilate (orange arrows) the contour (C) until we find that it is outside the EEL (red). More specifically, C is dilated until it satisfies:

- Condition 1: The value of the likelihood image at M points equidistantly distributed along the spline must be 0. This ensures that C is located outside the artery.
- Condition 2: At least 50% of the white pixels of the likelihood image must be inside C. This ensures that C stops outside the EEL.

B. Energies that drive the snake

A snake [10] is an energy-minimizing spline [11] guided by constraint and image forces. If we represent the position of a snake parametrically by $\mathbf{v}(s) = (x(s), y(s))$ where $s \in [0..1]$ we can then write the energy functional as:

$$E_{snake} = \int_0^1 E_{total}(\mathbf{v}(s))\, ds \qquad (4)$$
$$= \int_0^1 0.8 * E_s(\mathbf{v}(s)) + E_p(\mathbf{v}(s)) + E_{spl}(\mathbf{v}(s))\, ds$$

N break points (BP_i) control the snake.

As we can see in Equation (4), the energy functional is composed of three terms. A smoothness term (E_s), equivalent to the second derivative, prevents oscillations and sharp edges. A probability term (E_p) which is the maximum intensity value of the likelihood image found in a 5x5 window w_i centered in the point we are analyzing (BP_i). A spline term (E_{spl}), found by sampling the segment from BP_i to BP_{i-1} and the one from BP_i to BP_{i+1} and finding the average intensity value of all these sample pixels.

The optimal weights have been found experimentally.

V. RESULTS

To analyze the performance of our automatic algorithm, we consider the manual segmentation of the images as ground truth. We measure the error based on areas as described in Equation (5), where N_a is the number of pixels of the area found automatically and N_c the number of pixels of the manually segmented area.

$$Error = |N_a - N_c|/N_c * 100 \qquad (5)$$

Using this error measure, we analyze the impact of the size of the structuring element for erosion described in Equation(3). As shown in Figure 6, if parameter i is chosen in such a way that the number of white pixels left in the image is at least $PR = 60\%$ of the original white pixels, we have the best results (see Figure 7(a) for detailed results). Making the parameter variable improves by 10% the number of images correctly segmented, as opposed to using just a fixed erosion parameter of 50.
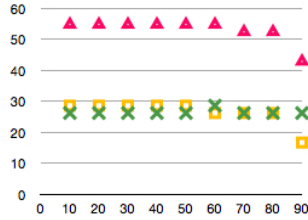
Fig. 6: X axis - Percentage PR for erosion calculation. Y axis - Cross: Error < 10%, Square: 10% < Error < 20%, Triangle: total error

We consider all the images with an error measure of 20% or less to be correct, as this does not affect the calculation of the fat in the artery, which is the ultimate objective.

A detailed analysis of the errors is shown in Figure 7(a). Note that "Error" just refers to the outer boundary, as in the blue and green cases the inner boundary is correctly segmented. Out of the 23.8% of images that had incorrect inner
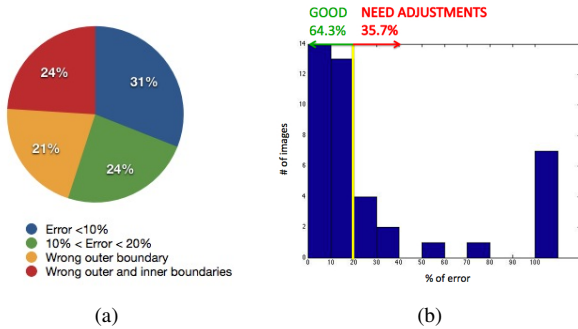

Fig. 7: (a) Detailed statistics of the error measures, (b) Final statistics of the error measures

boundaries, 90% can be fixed by adding seeds and 10% can be fixed by changing the threshold value. We take this into account to design our GUI which allows the user to easily:

- Observe the original and likelihood images
- Change the threshold and erosion parameters
- Recalculate the segmentation or change it manually

Therefore, with small adjustments using our GUI, we have 100% of the inner boundaries well-segmented.

In Figure 7(b), the statistics of the error measures including these small manual adjustments are presented. In conclusion, we obtain statistically 54.8% of automatically segmented outer boundaries and 10% more with minimal user interaction. From a user point of view, we can compute the average number of "mouse clicks" needed to segment an image as a measure of the workload of the user. 54.8% of the images are correctly analyzed with just one initial click (to choose the seed for region growing). All the images can be

analyzed with an average of 4.81 clicks.

VI. SUMMARY

In this paper, we have focused on the problem of segmenting arteries in multi-stain histology images. We have proposed an algorithm to automatically localize the artery and plaque within the image. Our approach is based on the creation of a statistical color model from training data, which allows us to obtain likelihood images. To find the outer boundary of the ROI in the likelihood images, we use snakes or active contours and minimize an error function which jointly optimizes a likelihood term, a smoothness term and a spline term. The initialization of the snakes is specially important to overcome the arteries shape differences and our solution is to use the inner boundary contour found by region growing. The overall algorithm achieves positive results that specially reduce the amount of time that the user spends analyzing the images. We obtain 64.3% of automatically segmented arteries. If we consider an average of 20 clicks to segment each image manually and compare it with our algorithm that allows the user to segment all images with an average of 4.81 clicks, we see that we reduce the user's work by 76%.

In the future, we plan to exploit the difference in texture between the media and the adventitia to make our algorithm more robust.

REFERENCES

1. Nair A, Kuban B.D, Tuzcu E.M, Schoenhagen P, Nissen S.E, Vince D.G. Coronary plaque classification with intravascular ultrasound radiofrequency data analysis. *Circulation.* 2002:2200-2206.
2. Meet F.J, Faber D.J, Perrée J, Pasterkamp G, Baraznji Sassoon D, Van Leeuwen T.G. Quantitative optical coherence tomography of arterial wall components *Lasers in Medical Science.* 2005;20:45-51.
3. Luo M. A colour management framework for medical imaging applications *Computerized Medical Imaging and Graphics.* ;30:357-361.
4. Herbin M, Venot A, Devaux J.Y, C Piette. Color quantitation through image processing in dermatology *IEEE Transactions on Medical Imaging.* 1990;9:262-269.
5. Pratt William K.. *Digital Image Processing 4th Edition.* John Wiley & Sons, Inc. 2007.
6. Niemisto A, Hu L, Yli-Harja O, Zhang W, Shmulevich I. Quantification of in vitro cell invasion through image analysis *Engineering in Medicine and Biology Society.* 2004;1:1703-1706.
7. Di Ruberto C, Dempster A, Khan S, Jarra B. Segmentation of blood images using morphological operators *15th International Conference on Pattern Recognition.* 2000;2:397-400.
8. Nedzved A, Ablameyko S, Pitas I. Morphological segmentation of histology cell images *Proceedings of the International Conference on Pattern Recognition.* 2000;1:500-503.
9. Maragos P, Schafer R.W, Butt M.A. *Mathematical Morphology and its applications to image and signal processing .* 1996.
10. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models *International Journal of Computer Vision.* 1988;1:321-331.
11. DeBoor C. *A Practical Guide to Splines.* Springer 1994.