HISTOLOGY IMAGE SEGMENTATION

Francisco J. Cisneros, Paula Cordero, Alejandro Figueroa, Juan Castellanos

Abstract: In this article, a technique for the segmentation of the components of histological images will be explained. To be able to do a study about the various microscopic components of the animal tissues and to reach to the histological images; first it is tried to be obtained some sections of the tissues and then; dye them in accordance with the different components which wanted to be studied. The image-analysis is a statistical work and most of the time, the data which is reached in the end, depends on the observer who is carrying out the study.

Keywords: Image Processing, Image Segmentation

ACM Classification Keywords: I.4.3 Image Processing and computer vision – Enhancement, J.3 Biology and Genetics

Introduction

Histology is a branch of science which studies the microscopic structures and the functions of the organic tissues. The actual means of technology; such as the microscopes, monitorized stages, slide loaders, cameras and PC; have improved the use of histology due to the incremental number of daily analyzed images. This increase allows having more precise data about the case-studies as we can have more reliable results. However, as the analysis of all sections of the tissues require a long time, only some sections of the sample are examined and the data from this sample is generalized to be used for statistical techniques (stereology). It has two faces; on the one hand, the analysis requires a long time while on the other - and also in the expanse of the other- it contains the problems of precision and reliability of the data. The possibility of the new automatic techniques in the image analysis addresses these problems and provides new advantages.

The automatic analysis of the images allows us not only to have reliable data but also it accelerates the acquisition process of the data. In addition to that, to be able by managing them digitally, it can also be reached to some other evaluation methods and data which are normally, expensive or not feasible; like the descriptors of textures or other descriptors which enables better analysis of the images. Thanks to those techniques, now we can carry out studies and reach conclusions which were not possible before.

The classic technique which is used to analyze the histological images is the stereology. The stereology is a method which enables the observers, to split three dimensional structures into two whenever they are parallel and equidistant. With the help of mathematical formulas which depend on the geometric probability, the results can be statistically significant.

Image analysis process requires a set of procedures; experimentation, processing and analysis procedures. First of all obtaining organic tissues to be analyzed, in order to get information, during the experimentation process these tissues are cut into thin slices to be observed under a microscopy. Before observation, slices are subjected to staining techniques to obtain the prepared pieces to be inspected and enhance contrast in the microscopic image. In observation process a set of images is taken through microcopy and then these images are processed using different techniques. In this context, it is necessary the segmentation technique's calibration in order to make plausible the distinction of the different areas to be identified in the image. Once the calibration is reached, images are segmented and descriptors are obtained from them (number of segmented regions, proportions, perimeter, etc). Image analysis and classification processes are carried out thanks to the information collected by these kinds of descriptors.
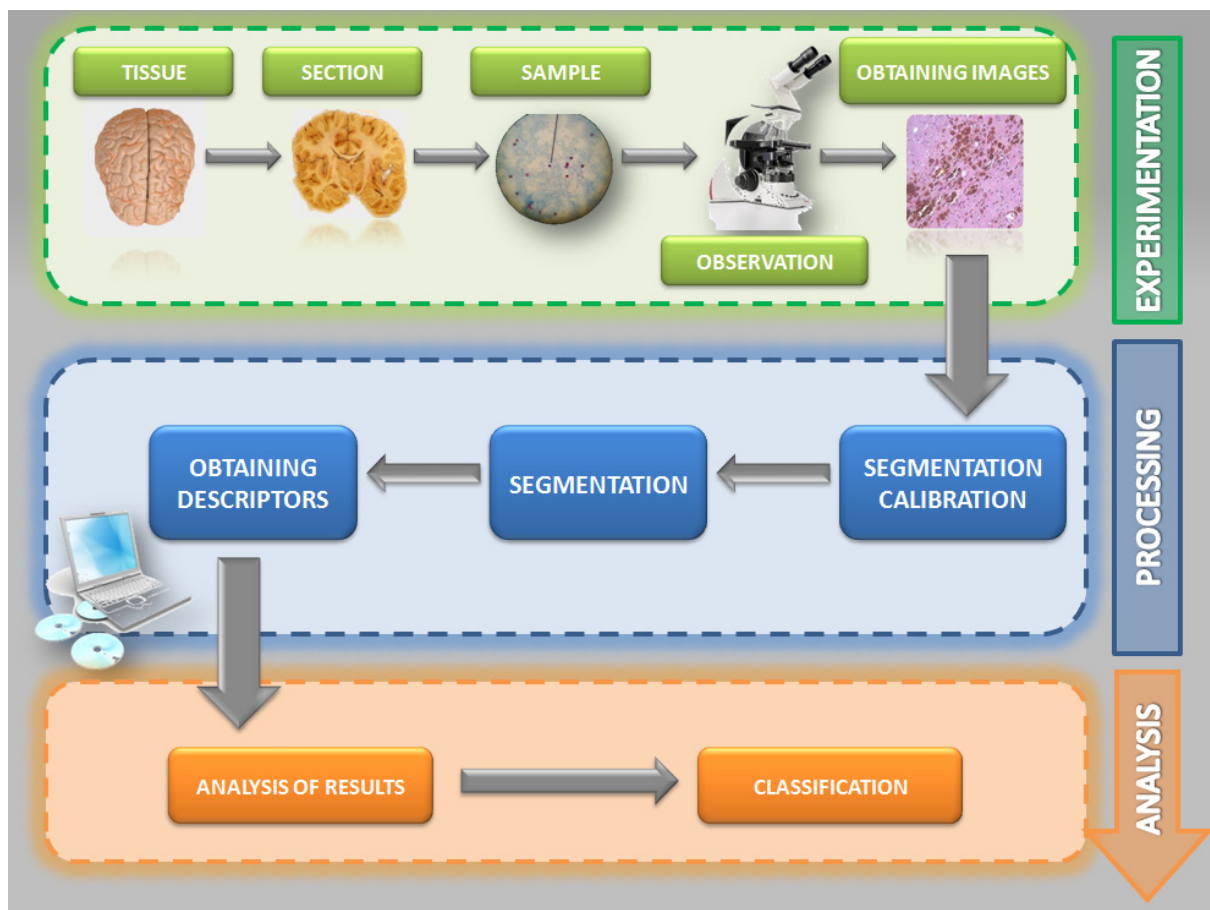

Figure 1.Experimentation, processing and analysis procedures of histology images.

Segmentation

Image segmentation is defined to be the process that subdivides an image into its constituent parts or objects   [R.C. Gonzalez and P.Wintz, 1987]. One of main challenges in image analysis is to identify and examine the tissue that   is often not present in the whole image of the study; the isolation of the desired tissue is carried out using segmentation processing. In addition, histology aims to analyze certain components of the same biological tissue and it is necessary to distinguish it within the image, this is the reason for using staining techniques for the different components to highlight structures for viewing, often with the aid of different microscopes. Identification of components in an image is easy for humans, but in order to get it using computational processes is necessary to apply segmentation techniques. After segmentation process you get an image divided into different compounds wanted to be analyzed, and it would be possible to be counted and analyzed with specific descriptors.

Nowadays there are many segmentation techniques and not all of them are valid for all environments. Due to a universal segmentation technique does not exist, it is necessary to use a concrete segmentation for each environment. Some main problems found during segmentation processes for the analysis of images are related to highly irregular image structures, inconsistent staining, non-uniform illumination, out-of-focus image components, and variability in the objects of interest. Segmentation techniques are divided primarily into these different topics*: Pixel-based-methods* as Threshold Segmentation, based on the categorization of pixels of an image according to a certain threshold (one pixel corresponds to a point on the digital image); Edge-based-methods as Edge Detection [Pal NR, 1993][Marr D, 1980], which consists of finding edges in the image in order to extract the closed polygons found; and finally *Region-based-methods* as Region Growing [MacQueen, 1967] based on group pixels with similar characteristics.

Segmentation techniques based on Edge Detection using minimum cost functions and information from the limits of the structures to   determine the   borders of homogeneous   regions. Some   of   these   kinds   of techniques use different filters to detect perimeters and they are usually based on the concept of gradient. When the result does not provide closed elements it is advisable to apply additional techniques to connect them like morphological   operators, for example. The   main problem   of   these techniques is   the   threshold used to determine the existence of an edge which   depends   heavily on image contrast and details. The   contrast, brightness and detail of the structures may be different from a tissue to another similar depending on the stain used   and the quality of it. This technique can   be interesting   for segmenting images   whose components are not clearly separated and parts of them overlap [J. Díaz, 2007].
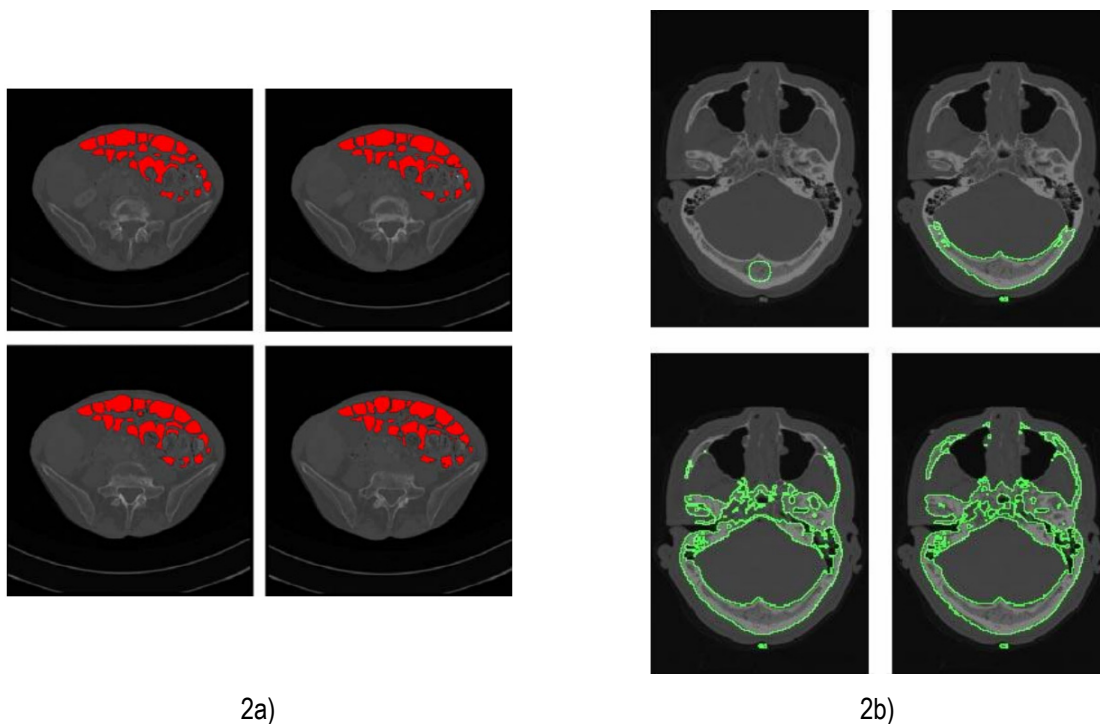
2a)   2b)

Figure 2. 2a) Region-growing segmentation from CT images. 2b) Automatic segmentation using a shape-based approach. Images from [J. Díaz, 2007].

Region-based-methods attempt to reconstruct different anatomical structures using a certain similarity criteria from the pixels of each selected area. Some of these segmentation techniques are known as Region-growing (figure 2a) approaches, starting from one or more initial regions called seeds which can be initialized manually or automatically using and heuristic method the initial regions are expanded to neighboring volumetric pixels following certain homogeneity criteria [D.L. Pham et al, 2000]. Region-growing techniques like Clustering have problems due to the great irregularity of the images and the variability of components to be segmented which sometimes cause an over-segmentation by the segmentation of not interesting components or interesting objects in too many regions.

Other methods of this family are Shape-based-approaches (figure 2b), which continuously keep a representation of the target element's shape. These shapes, also known as active shape models and active contour models, can be modified manually getting a better adaptation to the element or can be adjusted automatically according to the value of a certain energy function. Related to the application of shaped-based approaches it is found the well-known Level set-method, the main idea of this method is to evolve a curve towards the lowest potential of a cost function which must be defined imposing certain smoothness constraints based on Lagrange's method in order to parameterize the objective contour according to some sampling strategy and then evolve each element according to the image and internal terms [A. Tsai et al, 2003].

Histology image processing

Thanks to the experience in other histological images processing works, we observed that using thresholding techniques, we obtained very precise results in a really short computing time. However, with other methods like region growing techniques, we couldn't segment properly all the components we wanted to analyze. And also with edge detection techniques, in many times, we obtained an excessive segmentation plus additional processing to solve this problem.

Recall that histological images normally have the feature to tint the components to study in a specific color by being able to differentiate to the rest of the components by the color they present. Because of that, thresholding has turned out to be the most precise and efficient technique. This method examines the properties of each pixel in order to evaluate his color. The calibration of this technique is very simple because it just requires selecting a color sample of the component we want to analyze. This color sample turns out to be the same for every histological image accomplished with the same preparation (same dye, same tissue and same exposition time) .

Thresholding requires a measure to determine the type of component of the evaluated pixel. We propose the Mahalanobis distance to determine the kind of component being evaluated. In short, we measure the distance between the color of the pixel and the average color each one of the component we want to segment.

First the image is segmented using pixel-based techniques to classify pixels into component 1 and component 2 to be distinguished. Pixel-based techniques do not consider the spatial context but only decide on the basis of the color features at individual pixels. Pixel determination, component 1 vs. component 2, was based on the mahalanobis distance [Mahalanobis, 1936] to the chromatic gamma of the component 1 and setting an a priori threshold for the classification of first component pixels. Chromatic range was determined from seven component representative sampling. Measure is subjective and must be calibrated for each type of staining.

1. Select the color sample of the component we want to segment
2. Obtain the average color of the sample
3. For each pixel:
    1.1. Measure the Mahalanobis distance between the color of the pixel and the average color of the sample
    1.2. If this distance is lower than 0.15 the pixel is checked as a part of the component we want to segment
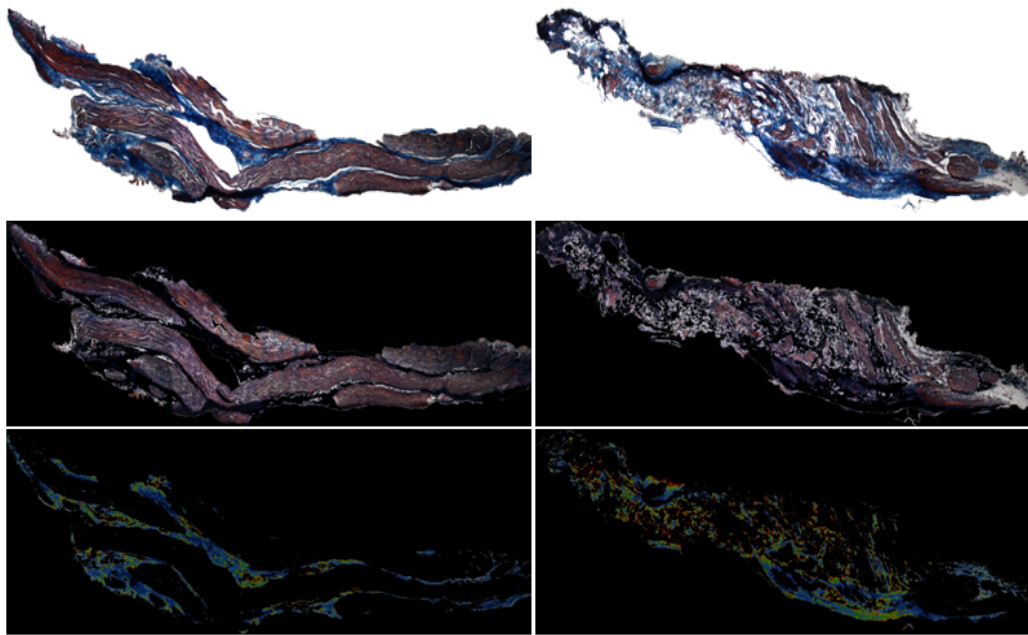4. It subdivides the image in each component, which keeps isolated

Figure 3. Histological neuroma image example [Herrera-Rincon, 2011]. Top: images of two neuromas (Prosthetic left and Amputated right). Middle: images after component 1 segmentation. Bottom: images after component 2 segmentation.

Figure 3 shows an example of the type of segmentation used in the article "Effects of Stimulation on the Peripheral Neuroprosthetic Nerves of Amputees: a Study Digital Image Processing" [Herrera-Rincon et al., 2011]. It was necessary a high level precision during the segmentation in order to make a correct count of component 1 percentage against component 2.

Experimentation

The following images were segmented using various techniques. In each example the result of applying the proposed segmentation is shown to be compared with others. After this results are discussed.

Figure 4 shows images obtained from tissues where it is intended to count cancer cells [Loukas et. al, 2003]. The technique used for segmentation is Laplacian of Gaussian (LoG) edge detector [Marr D, 1980]. For this technique we segmented only components (cells) in brown. As can be observed, for brown cells the results have been similar, though more problems that arises due to the proximity of all components to target.

The Figure 3 [Sertel et. al, 2008] show histological images of five major cytological components in the FL tissue: nuclei, cytoplasm, extra-cellular material, red blood cells (RBC) and background regions. Having nuclei and cytoplasm regions dyed with hues of blue and purple, extra-cellular material dyed with hues of pink and RBCs dyed with hues of red, H&E-stained FL images provides useful visual clues for segmentation. In addition to these components, there are also background regions that do not correspond to any tissue component. With this a priori knowledge on FL images, we performed the segmentation using K-means clustering algorithm [MacQueen, 1967] to identify these cytological components. With the proposed segmentation we obtain very similar results.

For this case we have selected the purple components (the nuclei). For each component is necessary to complete the complete process.
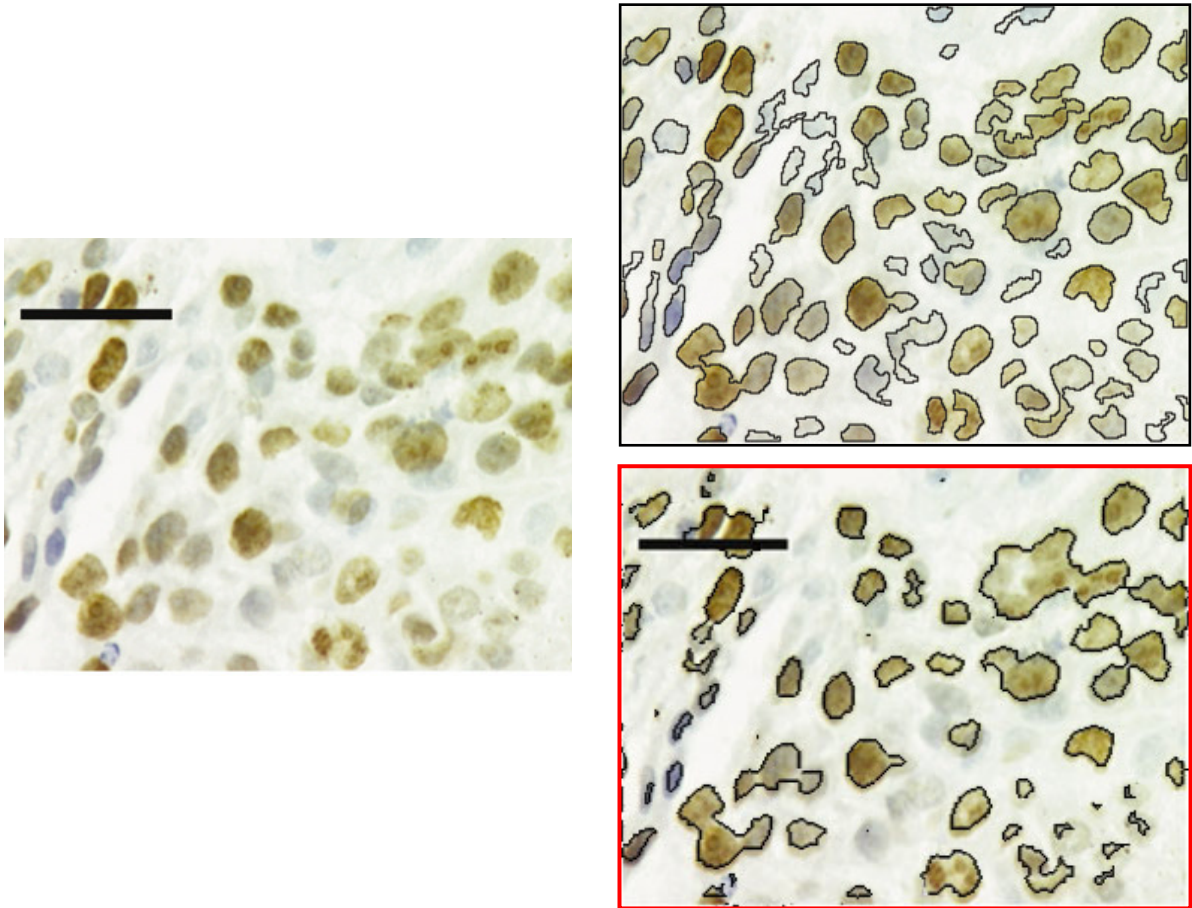

Figure 4. Example using Edge Detection [Loukas et. al, 2003]. Left, original image. Right, segmentation image (top, with LoG, bottom with our technique)

Local and Global Gaussian Mixture Models for Hematoxylin and Eosin Stained Histology Image Segmentation (LG-GMM) algorithm employs the unique characteristics of the H&E staining protocol: hematoxylin stains nuclei blue; eosin stains the cytoplasm pink, red blood cells red; air spaces are white. Therefore, we can obtain that nuclei are always represented by dark points. As it can be observed, in this case the results are practically the same, even more precise. This is due to good staining and spatial distance of the components to be detected.
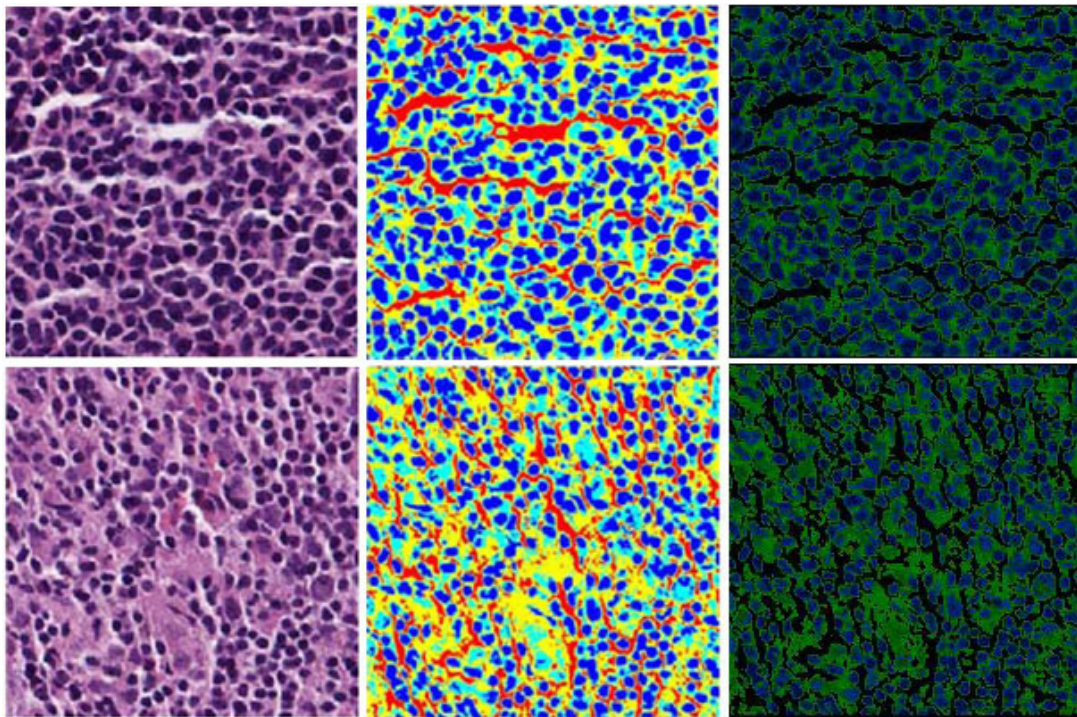
Figure 5. Example using K-means clustering [Sertel et. al, 2008]. Left column shows sample H&E-stained FL images. The corresponding segmentation with K-means results are shown in the center column. In these color labeled images, blue corresponds to nuclei, cyan to cytoplasm material and red and grey to background, and RBCs, respectively. The corresponding segmentation with us technique results is shown in the right column. In these, dark green to cytoplasm material, black to background, and green to RBCs.
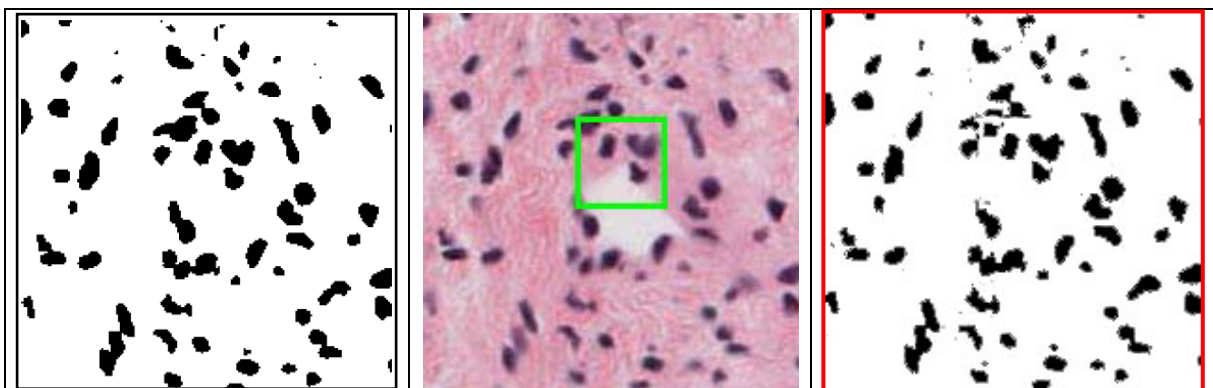

Figure 6. Example using LG-GMM segmentation [He et al, 2010] for nuclei detection example. First column shows the real image. The corresponding segmentation results are shown in the second column, the top correspond to LG-GMM segmentation, bottom correspond to the technique proposal in this article.

Conclusions

As shown in the results, the segmentation technique we propose has a very similar segmentation to that obtained by other techniques. The calibration of segmentation proposed is very simple, and the code's implementation is trivial in languages like MATLAB.

There are many image segmentation techniques, some of them very complex. Many of the papers published in recent years on the analysis of histological images have very sophisticated techniques for segmentation. This means spending more time in understanding, implementation and processing. The proposed technique provides results comparable and even better than those obtained with other techniques shown. Moreover, given the small number of parameters presented and easy understanding, it can be adapted quickly to any histological image and its processing speed is difficult to improve because of the small number of operations performed.

In this way we can conclude that the proposed segmentation technique has significant advantages for analysis of histological images. Especially in preparations with different dyes and whose elements have a considerable spatial separation.

References

[Herrera-Rincon et al., 2011] Celia Herrera-Rincon, Francisco Jose Cisneros, Juan Castellanos, and Fivos Panetsos. Effects of Neuroprosthetic Stimulation on the Peripheral Nerves of Amputees: a Digital Image Processing Study. 2011 5th International IEEE/EMBS Conference on Neural Engineering (NER). 2011

[He et al, 2010] Lei He, L. Rodney Long, Sameer Antani, George R. Thoma. Local and Global Gaussian Mixture Models for Hematoxylin and Eosin Stained Histology Image Segmentation. 10th International Conference on Hybrid Intelligent Systems. 978-1-4244-7364-9/10/$26.00 ©2010 IEEE pg 223-228. 2010

[Sertel et. al, 2008] Olcay Sertel, Jun Kong, Umit V. Catalyurek, Gerard Lozanski, Joel H. Saltz, Metin N. Gurcan. Histopathological Image Analysis Using Model-Based Intermediate Representations and Color Texture: Follicular Lymphoma Grading. J Sign Process Syst (2009) 55:169–183. DOI 10.1007/s11265-008-0201-y. 2008

[J. Díaz, 2007] Jose Díaz Iberri. Inspección Interactiva de Estructuras Anatómicas Tubulares. Master Thesis in Computer Sicinces. Universidad Politécnica de Cataluña. 2007

[A. Tsai et al, 2003] Andy Tsai et al. A Shape-Based Approach to the Segmentation of Medical Imagery Using Level Sets. IEEE Transactions on medical imaging. Vol. 22 No. 2. February, 2003

[Loukas et. al, 2003] Constantinos G. Loukas, George D. Wilson, Borivoj Vojnovic, and Alf Linney. An Image Analysis-Based Approach for Automated Counting of Cancer Cell Nuclei in Tissue Sections. Cytometry Part A 55A:30–42. 2003

[D.L. Pham et al, 2000] Dzung L. Pham et al. Current Methods in Medical Image Segmentation. Annual Review of Biomedical Engineering, volume 2, pp 315-337. 2000

[Pal NR, 1993] Pal NR, Pal SK. A review on image segmentation techniques. Pattern Recognition 1993;26:1277–1294. 1993

[R. C. Gonzalez and P. Wintz, 1987] R. C. Gonzalez and P. Wintz, Digital Image Processing. Addison-Wesley, Reading, Massachusetts. 1987

[Marr D, 1980] Marr D, Hildreth EC. Theory of edge detection. Proc R Soc London 1980;B207:187–217. 1980

[MacQueen, 1967] MacQueen, J. B. Some Methods for classification and Analysis of Multivariate Observations. In Proc. of 5th. Berkeley symp. on mathematical statistics and probability. (vol. 1, pp. 281–297) Berkeley, Univ. of California Press. 1967

[Mahalanobis, 1936] P. Mahalanobis, "On the generalized distance in statistics" Proc. Nat. Inst. Sci. India (Calcutta) , 2 (1936) pp. 49–55. 1936

Authors' Information

Francisco J. Cisneros – Natural Computing Group. Universidad Politécnica de Madrid, Boadilla del Monte, 28660 Madrid, Spain. email: kikocisneros@gmail.com

Paula Cordero – Natural Computing Group. Universidad Politécnica de Madrid, Boadilla del Monte, 28660 Madrid, Spain. email: pcordero@gcn.upm.es

Alejandro Figueroa – Natural Computing Group. Universidad Politécnica de Madrid, Boadilla del Monte, 28660 Madrid, Spain. email: afmeana@gmail.com

Juan Castellanos – Natural Computing Group. Universidad Politécnica de Madrid, Boadilla del Monte, 28660 Madrid, Spain. e-mail: jcastellanos@fi.upm.es