Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 1, January 2014, pg.226 – 232

RESEARCH ARTICLE

Segmentation of Nuclei in Cytological Images of Breast FNAC Sample: Case Study

Aditya P. Pise[1], Rushi Longadge[2], L. G. Malik[3]

[1] Department of Computer Science and Engineering, G. H. Raisoni Academy of Engineering and Technology, Nagpur, India
adityapise1@gmail.com

[2] Department of Computer Science and Engineering, G. H. Raisoni Academy of Engineering and Technology, Nagpur, India
rushi.longadge@gmail.com

[3] Department of Computer Science and Engineering, G. H. Raisoni College of Engineering, Nagpur, India
latesh.malik@raisoni.net

Abstract— ***Demand for increased robustness, better reliability and high automation of image segmentation algorithms is apparent in recent years. Precise diagnosis and prognosis is essential to reduce the high death rate. In this paper, the study of different methodologies of cytological image segmentation is proposed. The study includes the watershed algorithm and active contouring. One can also find here a description of de-noising and contrast enhancement techniques; because the raw image taken from camera mounted on microscope contain less information and noise. The study covers the different pre-segmentation processes, like Circular Hough Transform (CHT) for circle detection and nucleus localization method. Until now many segmentation algorithms were introduced but unfortunately those cannot be used directly for purpose of nuclei segmentation. From past few years' large efforts are taken to develop a fully automatic segmentation algorithm. Here, a group of modified versions of cytological image segmentation method adopted for fine needle biopsy images are presented. The discussion on common errors and possible future problems is also added.***

Keywords: CHT; Watershed; Active Contouring

I. Introduction

Breast cancer is the most frequently diagnosed cancer in women among the age group of 40 to 60. According to World Health Organization, worldwide each year 7.6 million deaths are caused by cancer, among which 502,000 are caused by breast cancer only [5]. With such a high rate, breast cancer also is one of the most lethal cancers. For many years, researchers have been trying to find the best way to treat breast cancer. Successful treatment is the key to reduce the high death rate. To successfully cure a patient from breast cancer we need to diagnose it as early as possible. Cancers in their early stages are susceptible to treatment while cancers in their most advanced stages are usually almost impossible to treat.

The most common diagnostic tools are fine needle aspiration biopsy (FNA) and mammography. Mammography is a non-invasive method, is most often used for screening purposes rather than for precise diagnosis. It allows physician to find possible position of micro-calcifications and other indicators in breast tissue. When a suspicious region is found, the patient is sent to a pathologist for a more precise diagnosis. This is when the FNA is taken. fine needle aspiration biopsy is an invasive method to extract a small sample of the questionable breast tissue that allows the pathologist to describe the type of cancer in detail. Using this method, pathologists can very adequately describe not only the type of the cancer but also its genealogy and malignancy. They can also foresee the course of cancer development by attributing a predictive factor to it. The determination of grade of malignancy is essential when predicting the progression of cancer.

In the last few decades a number of studies have been presented that aim at rendering the interpretation of cytological images more objective [24]. The success of these studies has been boosted by the increase in computational power of modern computer systems. This has allowed the automated performance of tasks which were otherwise performed manually. However, the majority of cytological and histological examinations are still performed manually, Here different experts opinion may differ because of various interpersonal biases [4].

In this paper, the study of different methods for segmentation of images obtained from microscope from breast FNAC is proposed. Our final goal is to isolate the nuclei of the cell from the rest cytoplasm. The Nucleus is very important structure within the call, because it is the place where breast malignancy can be observed [3]. Thus, much attention in the construction of the expert supporting diagnosis system has to be paid to the segmentation stage.

The main difficulty of the segmentation process is due to the incompleteness and uncertainty of the information contained in the image [1]. The imperfection of the data acquisition process in the form of noise, chromatic distortion and deformity of cytological material caused by its preparation additionally increases the problem complexity. The nature of image acquisition (3D to 2D transformation) and the method of scene illumination also affect the image luminance and sharpness. In many cases one must also deal with a low-cost CCD sensor whose quality and resolution capabilities are rather low.

Until now many segmentation methods have been proposed (Carlotto, 1987; Chen *et al.*, 1998; Kass *et al.*, 1987; Otsu, 1979; Su and Chou, 2001; Vincent and Soille, 1991) but, unfortunately, each of them introduces numerous additional problems and usually works in practice under given assumptions and/or needs the end-user's interaction/co-operation (Lee and Street, 2000; Street, 2000; Wolberg *et al.*, 1993; Zhou and Pycock, 1997).

II. CURRENT METHODOLOGIES

Image segmentation is perhaps the most studied area in computer vision, with numerous methods reported [10]. A segmentation method is usually designed taking into consideration the properties of a particular class of images. In this paper, a three-step segmentation method using the cytological images is discussed. The steps of our method are as follows.

2.1. Pre-processing

Though numerous enhancement algorithm are presented in literature [4, 25], but studies which are pertinent and suitable for this application are discussed in this section. Unlike the gray-scale image, color image does not carry significant information. So RGB color image should be converted to grayscale by removing blue and red chrominance components from the image defined in the YCbCr colorspace [3]. The luminosity component can be determined using:

$$Y = 0.299R + 0.587G + 0.114B. \quad (1)$$

Since the majority of images have low contrast; an enhancement technique is also needed to improve their quality(Fig.1).To reduce the complexity of algorithm, simple histogram processing with a linear transform (Contrast Stretching) of the levels of intensities of image can be used.

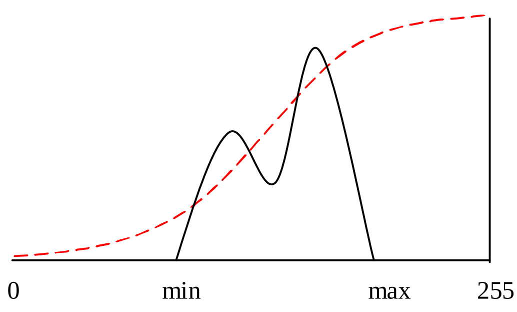

Fig.1. Typical histogram and a cumulated sum P (dashed line) [3]

The range of cutoff is defined by (*min*, *max*) points.

$$I_N = (I - Min)\frac{newMax - newMin}{Max - Min} + newMin \qquad (2)$$

newmax and *newmin* define a level in the histogram and in this approach, they are equal to 0.01 and 0.99, respectively, which means that 1% of pixels are saturated at low and 1% at high intensities of the input image[3]. By applying contrast correction in each color channel results in an image being better defined for later stages of the presented hybrid segmentation methods [1].

2.2. Presegmentation.

We could notice that the nuclei which we are willing to segment have the circular or elliptical shape (Fig.2).

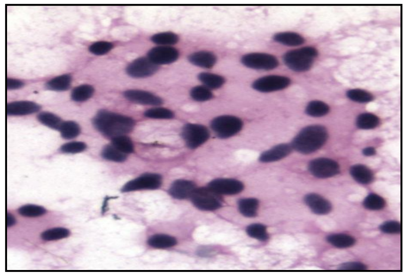

Fig.2. Microscopic Image of Breast FNAC sample

But regrettably, the detection of the ellipse which is described by two parameters *a* and *b* ($x=a\cos(\alpha)$, $y = b\sin(\alpha)$) and which can be additionally rotated is computationally expensive [6]. Hence, the detection of circles is much simpler with respect to the required computations because there is only one parameter, which is the radius *R*.

1) Method for Circle Detection: In this case, nuclei are our Region of Interest (ROI) which means our interest is to segment out the nuclei from rest image. And the most trusted and the most commonly used for circle detection in majority of papers is *Hough transform*.

In one of the methods there are two iteration of generalized Hough transform, one for obtaining knowledge of nuclei and another is for isolating the nuclei from themselves [25].These two iterations can be used as whole segmentation process. But this method is not worthy of use for segmenting all nuclei and also using two iteration of Hough transform is not efficient way of segmentation. Hence here only one iteration of Hough transform for detecting shape of nuclei is discussed.

*1.2) Hough Transform: O*ne of the most commonly used algorithms to recognize different shapes in an image is Hough Transform [8]. Hough Transform was introduced by Paul Hough in 1962 and patented by IBM. In 1972 Richard Duda and Peter Hart modified Hough Transform, which is used universally today under the name Generalized Hough Transform [9]. An extended form of General Hough Transform, Circular Hough Transform (CHT) [8] is used to detect circles. The edge detected from the canny edge detector forms the input to extract the circle using the Circular Hough Transform. In Circular Hough Transform, voting procedure is carried out in a parameter space. The local maxima in accumulator space obtained by voting procedure are used to compute the Hough Transform. Parameter space is defined by the parametric representation used to describe circles in the picture plane, which is given by equation (3). An accumulator is an array used to detect the existence of the circle in the Circular Hough Transform. Dimension of the accumulator is equal to the unknown parameters of the circle.
The equation of the circle in parametric form is given by

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \qquad (3)$$

Eq.3 implies that the accumulator space is three-dimensional (for three unknown parameters x0, y0 and r). This equation (3) defines the locus of points (x, y) centered on an origin (x0, y0) with radius r. Each edge point in figure 3a defines a set of circles in the accumulator space [8]. These circles are defined by all possible values of the radius and they are centered on the coordinates of the edge point. Figure 3b shows three circles defined by three edge points labeled 1, 2 and 3. These circles are defined for a given radius value. Each edge point defines circle for the other value of the radius. These edge points map to a cone of votes in the accumulator space (g), which is also called as feature space.

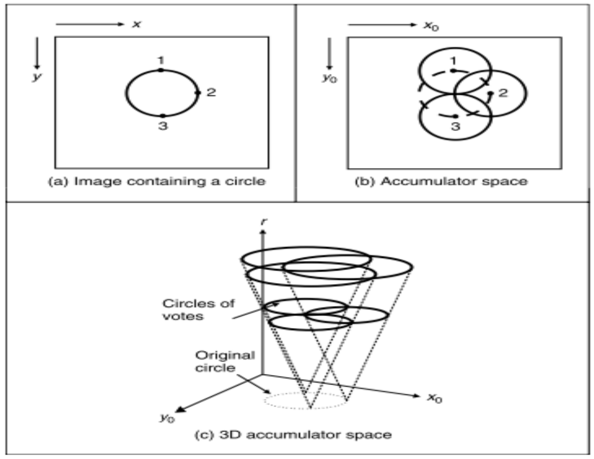

Fig3. Hough Transform illustration [14]

Figure 3c illustrates this accumulator. Points corresponding to x0, y0 and r, which have more votes, are considered to be a circle with center (x0, y0) and radius r, where *g* is a two-dimensional feature image. The feature space *g* can be created in many different ways. The study indicate that one should use gradient image as the feature indicating the occurrence or absence of the nucleus in a given fragment of the cytological image. The gradient image is a saturated sum of gradients estimated in eight directions on the grayscale image prepared in the preprocessing stage. The base gradients can be calculated using, e.g., Prewitt's or Sobel's mask methods [20] or their heavy or light versions (Fig. 4).

1	1	1
0	0	0
-1	-1	-1

1	2	1
0	0	0
-1	-2	-1

3	2	1
2	0	-2
-1	-2	-3

Fig.4. Gradient masks.

1.3. Threshold Selection : The values in the accumulator keeping as threshold by a given θ value, we can obtain a very good pre-segmentation mechanism with a lower threshold strategy [3]. Since the threshold value strongly depends on the database and the feature image *g*, the method can be used only as a pre-segmentation stage. A smaller value of the threshold causes fast removal of irrelevant information from the background, and what we achieve is a mask, which roughly defines the places where the nuclei and background are located.

1.3). Localizing Nucleus: The results obtained at the pre-segmentation stage can lead to the evaluation of an average background color. Such information can be used to model the nuclei as a color distance between the background and the objects, which fulfills the necessities of the lack of any color dependency in the imaged material. In our study we observe several distance metrics:

Manhattan's, Chebyshev's, the absolute hue value from the HSV color space, but the Euclidean distance is one which gives us visually the best results.

$$D_{euclid} = \sqrt{(I_r - B_r)^2 + \left(I_g - B_g\right)^2 + (I_b - B_b)^2} \quad (4)$$

Where B_c is the average background color estimated for the input image *I*. Since the modeling distance can vary in the local neighborhood, mostly because of camera sensor simplifications, a smoothing technique is needed to restructure the nuclei shape [3]. The smoothing operation in our approach relies on the fact that this sort of 2D signal can be modeled as a sum of sinusoids [14] with defined amplitudes, phase shifts and frequencies. Cutting off all low amplitude frequencies will results in a signal deprived of our problematic local noise effect. The frequency spectrum is determined using the discrete Fourier. After the frequency cutoff, the signal is reconstructed using the inverse of the discrete Fourier transform. And what finally achieve is a three dimensional modeled terrain where hills correspond to nuclei.

III. NUCLEI SEARCH

3.1. Multi-point version

There are two versions of this algorithm single-point and multi-point but this paper proposed efficient one which is multi-point algorithm. The algorithm that does not have such tight requirements concerning only one single marker per nucleus [13] can use a very simplified version of single point algorithm. In such cases use of only one population is discussed, i.e., the one searching for nuclei, and the fitness function is simply the terrain *height* in an individual position. The number of iterations of the algorithm can also be reduced, because only an approximate localization of nuclei is considered [12].

IV. IMAGE SEGMENTATION

4.1. Watershed

The watershed segmentation algorithm is inspired by natural observations, such as a rainy day in the mountains [20]. A given image can be defined as a terrain on which nuclei correspond to valleys (upside down the terrain modeled in previous steps). The terrain is flooded by rainwater and arising puddles start to turn into basins. When the water from one basin begins to pour away to another, a separating watershed is created. The flooding operation has to be stopped when the water level reaches a given threshold θ. The threshold should preferably be placed somewhere in the middle between the background and a nucleus localization point. In our approach the nuclei are flooded to the half of the altitude between the nucleus localization point and the average height of the background in the local neighborhood. Since the images we have to deal with are spot illuminated during the imaging operation (which results in a modeled terrain higher in the center of the image and much lower in the corners), this mechanism protects basins from being over-flooded and, in consequence, nuclei from being under-segmented [21].This method shows some primary errors like errors caused by fake circles created by spots of fat, that's why the Active contour method is more promising than the watershed method.

4.2. Active contours

The active contouring technique [25] can be considered as a more advanced region growing method [22]. The algorithm groups neighboring pixels when a given homogeneity and similarity criteria are met. All joined pixels create a segment whose boundary spreads in all directions until another segment is met or new candidates for joining introduce unacceptable errors. The algorithm is stopped when all pixels get labels, i.e., the object in the image is separated from the dbackground. The images may contain more than a single object per image. Additionally, the assumption of the project is that the segmentation process has to be fully automatic (there is no human operator which manually initializes the method). These two factors force us to modify the algorithm to meet the stated requirements. Thus, the algorithm, which in this case, it is based on the fast marching method (FMM) [17], must have a multilevel extension [14] and the seeding process which has to be done without the end-user's interaction. The modified multilevel FMM algorithm is very stable and robust to initialization errors. Visually, segmentation quality is promising and yields a good detection of even small objects [3].

V. DISCUSSION

Analysis drawn from comparative study of each of the methods is shown in Table 1.This paper proposed different methods, and every method is subjective to that particular to application. The solution provided by many techniques in segmentation is helpful but not enough. The consideration of which technique is best for handling a problem of image segmentation highly depends upon the nature of images used for experiment. In each case we observed that the problem of connected or overlapping nuclei is still there because it has same intensity and structure and it is impossible to segment each nuclei. May be in future this can be removed by advanced methods. In this discussion we observed that the one method (Maciej Hrebien´ , Piotr Stec`,2008) gives visually best result as segmentation concern(Fig.5a,5b). Each separate color ellipse shaped spot in figure 5b is a separate nucleus.

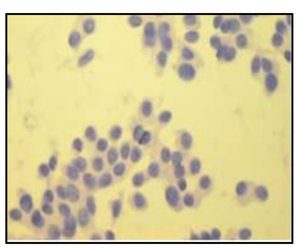

Fig.5a Orignal Image

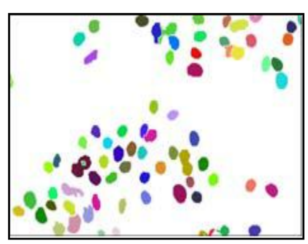

Fig.5b.Segmented nuclei

[3]

TABLE 1. COMPARATIVE STUDY

Sr. No.	Methods/ Algorithms	Advantages	Dis-advantages
1	Kyoung-Mi lee,street, (2000)	Improvement in time needed to perform template - matching.	End-users interaction is needed.
2	Wolberg W. Street W. and Mangasarian O. (1993).	The system yields more accurate than any other methods discussed	End-users interaction is needed. And is also not efficient
3	Zhou P. and Pycock D. (1997).	The Gaussian model and Bayesian distance metric is cost effective.	Performance of this algo. is Comparatively less.
4	Marek Brejl,M. Sonka, (2000)	All necessary information about nuclei derived from training data set.	Computational complexity and computa-tional time is large
5	Street W. (2000) , (Xcyt System)	Provides remote predictive analysis for breast cancer diagnosis and prognosis.	End-users interaction is needed.
6	Maciej,Piotr,,Tomasz,Andrzej(20008)	The segmentation process is fully automatic.	Problem of connected nuclei still there.

VI. **CONCLUSION**

The Maciej,Piotr,,Tomasz,Andrzej(20008) method discussed in this paper reduces the exercise time required for image scaling and extract nuclear boundaries. Summarizing the presented solutions, promise and give a good base for further research in the area of cytological image segmentation. Additionally, all preparation steps including pre-segmentation and the automatic nuclei localization stage can be reused with other segmentation algorithms which need such information .In this approach each and every process could be done automatically hence our ultimate aim can be achieved.

ACKNOWLEDGMENT

We would like to thank the following people for their help Dr. Hrushikesh T. Garud, P. Deshmukh, and Dr. Y. Rajwade, A. Parekh and Dr. Shivraj Suryawanshi.

REFERENCES

[1] Marek Brejl and Milan Sonka,"*Object Localization and Border Detection Criteria Design in Edge-Based Image Segmentation: Automated Learning from Examples*" Ieee Transactions On Medical Imaging, Vol. 19, No. 10, October 2000.

[2] C.Ortiz De Solo´, Rzano,E., Garci´A Rodriguez,A. Jones,D. Pinkel, J. W. Gray, D. Sudar & S. J. Lockett, "*Segmentation of confocal microscope images of cell nuclei in thick tissue sections*" *Journal of Microscopy, Vol. 193, Pt 3, March 1999, pp. 212–226.*

[3] Maciej Hrebien´ , Piotr Stec´ , Tomasz Nieczkowski, Andrzej Obuchowicz, "*Segmentation Of Breast Cancer Fine Needle Biopsy Cytological Images*" Int. J. Appl. Math. Comput. Sci., 2008, Vol. 18, No. 2, 159–170.

[4] Theodorus Mouroutis, Stephen J.Roberts, Anil A.Bharath, "*Robust Cell Nuclei Segmentation Using Statistical Modelling*" Bioimaging,1998,citeseer.Pg.No 79—91,vol. 6.

[5] Łukasz Jelen´ , Thomas Fevens, Adam Krzyz˙ Ak, "*Classification Of Breast Cancer Malignancy Using Cytological Images Of Fine Needle Aspiration Biopsies*" Int. J. Appl. Math. Comput. Sci, 2008, Vol.18,No.1,75–83.

[6] Marcin Smereka, Ignacy Dule¸Ba, "Circular Object Detection Using A Modified Hough Transform" Int. J. Appl. Math. Comput. Sci., 2008, Vol. 18, No. 1, 85–91.

[7] Chunming Li, Chenyang Xu, Changfeng Gui, and Martin D. Fox, "*Level Set Evolution Without Re-initialization: A New Variational Formulation*" Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 1063-6919/05 $20.00 © 2005 IEEE.

[8] F. Yan, X. Shao, G. Li, Z. Sun, and Z. Yang, "Edge detection of tank level IR imaging based on the auto-adaptive double-threshold canny operator," *Intell. Inform. Technology Applicat. Research Assoc.*, vol. 3, pp. 366 - 370, Dec. 2008.

[9] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Commun. the ACM*, vol. 15, pp. 11-15, Jan. 1972.

[10] L. Xua, M. Jackowskia, A. Goshtasbya,*, D. Rosemanb, S. Binesb, C. Yuc, A. Dhawand, A. Huntleye, "Segmentation of skin cancer images" Image and Vision Computing 17 (1999) 65–74,ELSEVIER.

[11] W. N. Street, "*Xcyt: A System for Remote Cytological Diagnosis and Prognosis of Breast Cancer*".

[12] Borovec, Jiří. (2013). Fully automatic segmentation of stained histological cuts. In L. Husník (Ed.), 17th International Student Conference on Electrical Engineering (pp. 1–7). Prague: CTU in Prague.

[13] Arabas J.. "*Lectures on Evolutionary Algorithms*", WNT, Warsaw (in Polish) (2004).

[14] Ballard D. "*Generalizing the Hough transform to detect arbitrary shapes*", Pattern Recognition **13**(2): 111-122 (1981).

[15] Duda R. and Hart P. "*Use of the Hough transformation to detect lines and curves in picture*", Communications of Association for Computing Machinery **15**(1): 11-15 (1972).

[16] Marciniak A., Obuchowicz A., Monczak R. and Kołodzi´nski M. "*Cytomorphometry of fine needle biopsy material from the breast cancer*", *Proceedings of the 4th International Conference on Computer Recognition Systems CORES'05,* Springer, Rydzyna, Poland, pp. 603-609. (2005).

[17] Otsu N. "*A threshold selection method from grey-level histograms*", *IEEE Transactions on Systems, Man and Cybernetics* **9**(1): 62-66 (1979).

[18] Pratt W. *Digital Image Processing*, Wiley, New York. (2001).

[19] Russ J. *The Image Processing Handbook*, CRC Press, Boca Raton, FL (1999).

[20] Gonzalez R. and Woods R. (2002). *Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ.

[21] en.wikipedia.org/wiki/YCbCr.

[23] HGalera_Davidson,R.Gonzalez-Campora,A.Mora-Marin "*Cryptographic DNA measurement in medullary thyroid carcinoma*". Cancer,65:2255 2260,1990.

[24] R.Bowman and G.Angilini, "Grading of astrocytomas using neural network and linear discriminant functions based on morphometric mesurment collacted using static image cytometry". America Jornal of clinical pathology,102:502,1995.

[25] Setiono R."*Extracting rules from pruned neural networks for breast cancer diagnosi", Artificial Intelligence in Medicine* **8** (1): 37-51. (1996)

Authors

1. Aditya P. Pise, Mtech Student, Department of Computer Science and Engineering, G. H. Raisoni Academy of Engineering and Technology, Nagpur, India.

2. Rushi Longadge, Assistant Professor, Department of Computer Science and Engineering, G. H. Raisoni Academy of Engineering and Technology, Nagpur, India.

3. Mrs. L. G. Malik, Professor and HoD, Department of Computer Science and Engineering, G. H. Raisoni College of Engineering, Nagpur, India.

