Adaptive Local Thresholding for Detection of Nuclei in Diversity Stained Cytology Images

Neerad Phansalkar† Sumit More Ashish Sabale Madhuri Joshi
Department of Electronics and Telecommunication
College of Engineering Pune,
Pune, India.
†phansalkarnd06.extc@coep.ac.in

***Abstract*—Accurate cell nucleus segmentation is necessary for automated cytological image analysis. Thresholding is a crucial step in segmentation. The accuracy of segmentation depends on the accuracy of thresholding. In this paper we propose a new method for thresholding of photomicrographs of diversly stained cytology smears. To account for the different stains, we use different color spaces. A new local thresholding scheme is developed to solve the problem of nonuniform staining. Finally, the results obtained from the new method are compared with those of some of the existing thresholding methods, clearly showing the improvement achieved.**

Keywords— cytology; image segmentation; local thresholding; color spaces

I. INTRODUCTION

Segmentation of cell nuclei is a fundamental subject of quantitative analysis of cytological and histological images. As many of the characteristics of a cell are contained in the nucleus, the separation of nuclei from the background is an important part of segmentation for this kind of cell images. This separation of the nuclei from the background is done through either global or local thresholding.

In the analysis of cytology images, determination of threshold is a particularly difficult problem to solve, because of the diversity of structures contained in these images and the intensity variations of the background and foreground caused by uneven staining. Also, different stains are used for staining cytology smears, which makes the design of a generalized segmentation scheme difficult.

In view of the above problems, some thresholding methods have been proposed in the past [1]. The optimum global thresholding scheme proposed by Otsu may well be applied to the case of cytology images [2]. However, being a global scheme, this method cannot account for the local variations in the intensity caused by nonuniform staining. Thus, due to nonuniform staining, a global thresholding scheme is inappropriate for the case of cytology images. The methods of local thresholding are quite well developed in the field of document analysis, unlike the field of cytological image analysis. Numerous methods about the former can be found in literature [3]. Thus, there is a need to develop a thresholding method for the later, for accurate segmentation. Starting from a local thresholding method proposed by Sauvola and Pietikainen in the field of document image analysis [4], we develop a method to suit the needs of our problem.

Further, in view of a generalized system for analysis of cytology images, the developed method should work irrespective of the stain used. To account for the different stains, the images are first preprocessed using different color spaces.

II. MATERIALS AND METHODS

A. Image Specifications and Acquisition

The images used in this study are cytology smears of cancerous as well as noncancerous tissues of various organs in the human body. Cancerous specimens were specifically taken because the malignant nuclei have some irregularities, and the developed method should work irrespective of these irregularities. The magnification of all the images is 400. The smears are stained in either H & E stain, Pap stain, or Giemsa stain.

B. Stains and Color Spaces

In this section, we discuss the preprocessing step of the algorithm. However, before going into the details of the preprocessing, the different stains used in cytology and their characteristics are briefly presented.

Pathologists use different stains for staining different cytological smears (specimens). This is because some tissues are stained better in some stains than others. The three major stains used in cytology are Hematoxylin and Eosin (H & E) stain, Papanicolaou (Pap) stain and Giemsa stain. However, whatever be the stain, it is observed that the nuclei are distinguishable from the background because of a difference in either intensity, or color, or both. To be specific, the nuclei and background in the H & E stain image are well separated from each other in color, and in Giemsa and Pap stain images, they differ in intensities. We use the RGB and L*a*b* color spaces, for the reasons explained below. The L*a*b* color space image can be found from the RGB color space image using the conversions presented in [5].

The RGB image is converted to grayscale and L*a*b* image is converted to a Euclidean distance image, determined using L*, a*, b* as the coordinates [5]:

$$I_{RGB} = 0.3 * R + 0.59 * G + 0.11 * B \quad (1)$$

$$I_{Lab} = \sqrt{(L^*)^2 + (a^*)^2 + (b^*)^2} \quad (2)$$

where I_{RGB} and I_{Lab} are the monochrome images obtained from RGB and L*a*b* images respectively. These monochrome images are then normalized. The grayscale image obtained from the RGB color space provides better contrast between the nuclei and the background, when

978-1-4244-9799-7/11/$26.00 ©2011 IEEE

they differ in intensities. Likewise, the monochrome image obtained from the L*a*b* color space provides a better contrast between the background and the nuclei when they differ in color.

C. Thresholding

Based on the technique used in the method proposed by Sauvola and Pietikainen for local thresholding of document images [4], we developed a method suitable for our problem. In their method, the local threshold T(x,y) is found from the mean m(x,y) and standard deviation s(x,y) of the pixel intensities in a w × w window centered around the pixel (x, y) such that,

$$T(x,y) = m(x,y)\left[1 + k\left(\frac{s(x,y)}{R} - 1\right)\right] \tag{3}$$

where R is the dynamic range of standard deviation which is equal to 0.5 for a normalized image and k is a constant which takes values in the range [0.2,0.5] (in our experiment k = 0.25 gave best results). The local mean m(x,y) and standard deviation s(x,y) adapt the value of the local threshold according to the contrast in the local neighborhood of the pixel. When there is high contrast in some region of the image, s(x,y) ≈ R such that T(x,y) ≈ m(x,y). However, when the contrast in the local neighborhood is quite low, the value of the threshold goes below the mean, thereby successfully removing the relatively dark regions of the background.

However, unlike document images, cytological images have low contrast dark regions also (representing foreground), which correspond to clusters of nuclei, or even large sized individual nuclei. When the window is over such a region, the mean will be quite low and so will the standard deviation. Hence, the threshold found out by this method will be very low. Due to this, many of the pixels which actually correspond to nuclei will be rejected. Hence, we need to devise a method which will work the same as Savoula's method for high values of the local mean; however, for lower values of local mean, the threshold determined should be much higher than that determined by Sauvola's method, even higher than the local mean. This can be done by computing the threshold as,

$$T(x,y) = m(x,y)\left[1 + pe^{-q\cdot m(x,y)} + k\left(\frac{s(x,y)}{R} - 1\right)\right] \tag{4}$$

where p and q are constants. The value of q is so chosen that above a particular value of the local mean, the exponential term becomes negligible and the equation behaves just like Sauvola's equation. For choosing this value, we first stretch the histograms of the normalized monochrome images obtained from RGB and L*a*b* between 0 and 1 (i.e. the minimum value in the image is set to zero and maximum to 1). When this is done, we observed that the values of the pixels corresponding to nuclei in both images lie below 0.4. Thus when the value of the mean is less than 0.4, the exponential term should come into picture. We know that an exponential of the form given above reaches approximately 98% of its final value when the value of the exponent becomes 4. Here, the exponential term is required to go down to zero when the mean is 0.4, i.e.

$$q \cdot 0.4 = 4 \tag{5}$$

$$q = 10. \tag{6}$$

The constant p decides the magnitude to which the exponential term will affect the threshold. For very low values of p (i.e. $0 \le p \le 1$), the method behaves almost like Sauvola's method. For very high values of p ($p > 5$), the threshold becomes too high and too many background pixels are classified as foreground. In our experiment, p = 3 gave the best results. Fig. 1 shows the variation of the threshold against the local mean for both proposed and Sauvola's methods, with the values of constants as given above. The graph has been plotted for the minimum value of standard deviation i.e. zero.

As can be seen from the graph, the value of the threshold for low values of mean for the proposed algorithm is significantly higher than that for Sauvola's algorithm. Thus, now when the window is over a nuclear cluster (or a large nucleus), the value of the threshold is much higher than that obtained by Sauvola's method, and we get accurate binarization results. At high values of the local mean, however, the two curves coincide.

Further, using the concept of integral images [6], a fast algorithm can be implemented for this method on the lines given in [7]. In order to speed up the process further, we first apply a global threshold to the image, and only the pixels selected in this step as foreground are applied the local threshold. Such a global threshold should not reject a single pixel that corresponds to a nucleus (it may, however, select nonnuclear pixels, which will be removed during local thresholding). Such a threshold is obtained using Otsu's method.

The thresholding is applied to both RGB and L*a*b* monochrome images. Thus, after the thresholding step, we get two binary images, BW_{RGB} and BW_{Lab}, for the two monochrome images respectively. In order to find the most suitable binarization result irrespective of the stain, these two images are ORed (assuming that the black pixels in the images represent the nuclear matter) to get the final binary image, BW:

$$BW = (BW_{RGB}) \mid (BW_{Lab}). \tag{7}$$

where '|' denotes the pixelwise OR operator.

III. Results and Discussion

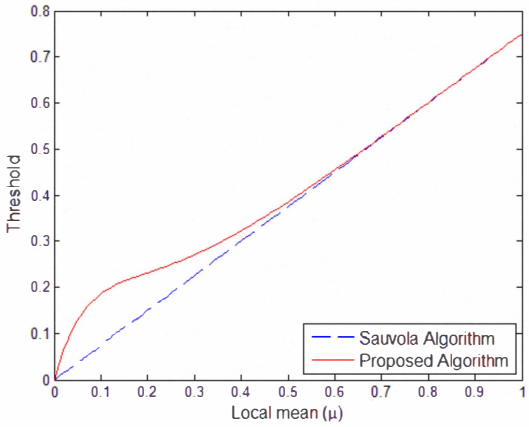


Figure 1: Variation of threshold against local mean in Sauvola's method and the proposed method for minimum and maximum values of standard deviation.

The results of application of the proposed method, Sauvola's method and Otsu's method are shown in Fig. 2 and Fig. 3. A window size of 40×40 was used for both Sauvola's and proposed methods. All the three methods are applied to RGB as well as L*a*b* monochrome images, and then the resulting two binary images are ORed as explained above to give the results shown in the figures. As we can see, the results of the proposed method are more accurate than any of the other two methods. Specifically, Otsu's method results in too many background pixels being classified as foreground. Conversely, Sauvola's method results in too many foreground pixels being rejected, due to dense clusters or large sizes of the nuclei. The proposed method gives the most accurate classification of background and foreground. Thus, in the results of the proposed method, the nuclei are better defined, and very few nuclei are lost (either partially or totally) unlike the other methods.

IV. Conclusion

The problem of segmentation of nuclei in analysis of cytology images was investigated. A method was proposed to make the analysis independent of the stain used. Further, a new local thresholding method was proposed to address the problem of nonuniform staining. The developed thresholding method was shown to perform better than two of the existing and most widely used methods.

Acknowledgment

The authors would like to thank Alluri Sitaram Raju Academy of Medical Sciences, Eluru, India, for providing the valuable database of photomicrographs of cytology slides.

References

[1] C. Garbay, "Synthetic Image Structure Representation and Processing: A Discussion of Some Segmentation Methods in Cytology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 2, pp. 140-146, Mar. 1986.

[2] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62-66, Jan 1979.

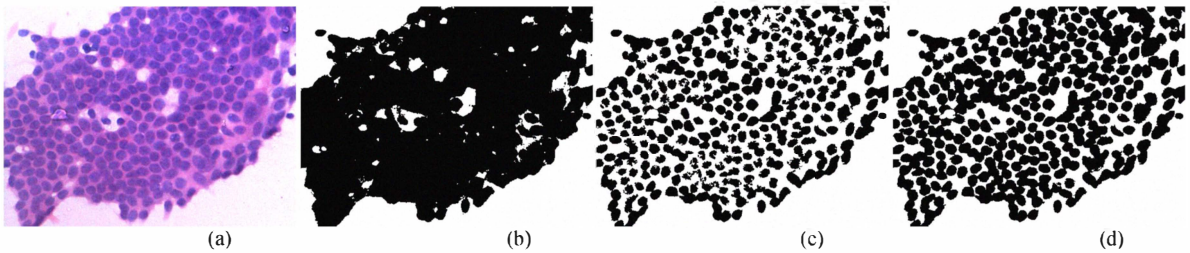
[3] M. Sezin, B. Sankur, "Survey Over Image Thresholding Techniques and Quantitative Performance Evaluation," *J. Electronic Imaging,* vol. 13, no. 1, pp. 146-165, Jan. 2004.

[4] J. Sauvola, M. Pietikainen, "Adaptive Document Image Binarization," *Pattern Recognition,* vol. 33, pp. 225-236, 2000.

[5] H. Cheng, X. Jiang, Y. Sun, J. Wang, "Color image segmentation: advances and prospects," *Pattern Recognition,* vol. 34, no. 12, pp. 2259-2281, Dec. 2001.

[6] P. Viola, M. J. Jones, "Robust real-time face detection," *Int. J. Computer Vision,* vol. 57, no. 2, pp. 137-154, 2004.

[7] F. Shafait, D. Keysers, T. M. Breuel, "Efficient Implementation of Local Adaptive Thresholding Techniques Using Integral Images," in *proc. SPIE Document Recognition and Retrieval XV,* San Jose, CA, Jan. 2008, pp. 26-31.



(a) (b) (c) (d)

Figure 2: (a) Photomicrograph of a cytology smear, (b) result of thresholding using Otsu's global thresholding method, (c) result of thresholding using Sauvola's method, (d) result of thresholding using the proposed method. Due to dense clusters of nuclei, a lot of nuclear pixels are rejected by Sauvola's method, while our method gives a very accurate result.

(a) (b) (c) (d)

Figure 3: (a) Photomicrograph of a cytology smear, (b) result of thresholding using Otsu's global thresholding method, (c) result of thresholding using Sauvola's method, (d) result of thresholding using the proposed method. Due to large sizes of the nuclei, Sauvola's method leaves out the inner parts of the large nuclei. Our method gives a much better result.