cytology, image processing, segmentation

Maciej HREBIEŃ, Piotr STEĆ

THE HOUGH TRANSFORM AND ACTIVE CONTOURS IN SEGMENTATION OF CYTOLOGICAL IMAGES

This paper describes a hybrid segmentation method of cytological images. The analysis includes the Hough transform and the Active Contours method. One can also find here a short description of image pre-processing used in our approach and preliminary experimental results collected on a prepared benchmark database.

1. INTRODUCTION

Automatic cancer diagnosis systems have a very long history and are still under dynamic research by many university centres and commercial institutions [3]. Since breast cancer is becoming most common disease of the female people of today, many efforts were and still should be done to increase its detectability.

A nucleus of the cell is the place where a breast cancer malignancy can be observed. Therefore, it is crucial for any camera-based automatic diagnosis system to separate cells from the rest of the image. Until now, many segmentation methods were proposed [2, 6, 7] but unfortunately each of them introduces different kinds of additional problems. Because many cytological projects assume full automation and real-time operation with high degree of efficacy, the method deprived drawbacks of already known approaches have to be constructed.

In this paper a hybrid method combining the Hough transform with the Active Contour method for cytological image segmentation is presented. The paper also presents pre-processing used in our approach and preliminary experimental results collected on a prepared benchmark database.

2. IMAGE PRE-SEGMENTATION

2.1. IMAGE PRE-PROCESSING

The quantity of information contained in a colour image is surplus at the early stage of image processing. The colour components do not carry as important information as luminosity so they can

Institute of Control and Computation Engineering
University of Zielona Góra, ul. Podgórna 50, 65-246 Zielona Góra
Emails: {m.hrebien, p.stec}@issi.uz.zgora.pl

be removed to reduce processing complexity. An RGB colour image can be converted to greyscale by calculating a luminance value in the same way as it is calculated for YCbCr colour space [6].

Since a great deal of images have a low contrast, an enhancement technique is needed to improve theirs quality. In our research we use a simple histogram processing with the linear transform of images levels of intensities, namely a cumulated sum approach [7].

2.2. THE HOUGH TRANSFORM FOR CIRCLE DETECTION

It can be observed that the cells we have to segment have elliptical shape. Unfortunately, detection of ellipses, which are described by two parameters *a* and *b* ($x = a\cos\alpha$, $y = b\sin\alpha$) and which can be additionally rotated, is computationally expensive. The shape of ellipses can be approximated by a given number of circles. Detection of circles is much simpler, in the sense of required computations, because we have only one parameter, that is radius *R*.

The Hough transform [12, 13] can be easily adopted for the purpose of circle detection. The transform in the discrete space could be defined by:

$$HT(R,\hat{i},\hat{j}) = \sum_{i=\hat{i}-R}^{\hat{i}+R} \sum_{j=\hat{j}-R}^{\hat{j}+R} g(i,j)\delta\left((i-\hat{i})^2 + (j-\hat{j})^2 - R^2\right), \tag{1}$$

where *g* is a two dimensional feature image and δ is the Kronecker's delta (unit answer at zero) defining sum only over the circle. The *HT* plays the role of accumulator which accumulates level of similarity of feature image *g* to circle placed at the $(\hat{i},\hat{j})$ position and defined by the radius *R*.

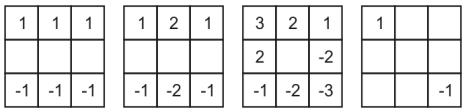


Fig. 1. Gradient masks used in our experiments.

The feature image *g* can be created in many different ways. In our approach we use gradient image as the feature that indicates cells occurrence or absence in a given fragment of cytological image. The gradient image is a saturated sum of base gradients estimated in eight directions. The base gradients can be calculated using e.g. Prewitt's, Sobel's mask methods [11] or their heavy or light versions (Fig. 1).

2.3. SUMMARY AND EXEMPLARY RESULTS

The main problem of the presented segmentation method is a proper selection of *HT* accumulator threshold value for which we suspect existence of a circle (cell) at a given image position (Fig. 2). Since the threshold (level of coverage) strongly depends on the database and used feature image *g*, the method can be used as a pre-segmentation mechanism for more computationally expensive segmentation algorithms. Thus the threshold can be given a smaller value and, because the method is relatively fast, the non-important information from background can be quite quickly re-

moved. Additionally, the method can be easily decomposed to perform even more effectively on SIMD organized machines and parallelized for multithreaded systems.

Exemplary results of image pre-segmentation for different *HT* threshold values are shown in Fig. 3.

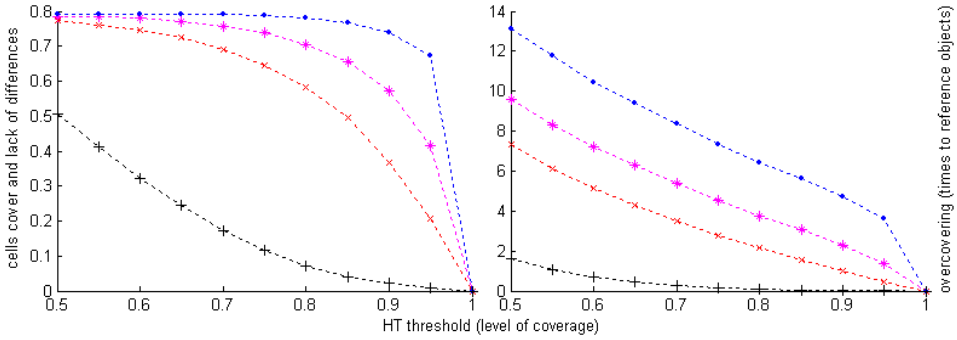
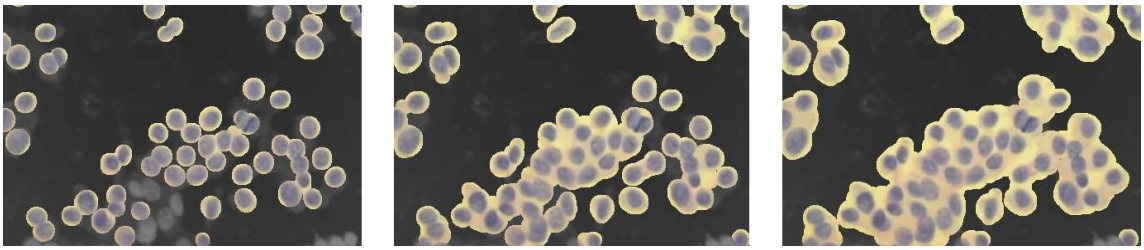


Fig. 2. Charts illustrating: influence of *HT* threshold value on cell cover and lack of differences (left) and overcovering (right) for Prewitt (×), Sobel (*), heavy (•) and light (+) base gradient masks (experiments performed on a randomly selected 346 element Zielona Góra's ONKOMED [5] cytological benchmark database for radiuses in the 4-21 range).



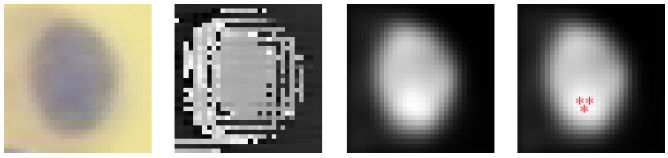
(a) (b) (c)

Fig. 3. Exemplary results of image pre-segmentation for: (a) high, (b) middle and (c) low *HT* threshold values.

3. IMAGE SEGMENTATION

3.1. CELLS LOCALIZATION

The results obtained from the pre-segmentation mechanism can lead us to the estimation of average background colour. This information can be used to model the cells as a colour distance between background and objects. In our approach we use HSV colour space and the distance is defined as the absolute Hue value difference between background and objects. Because the distance can vary in local neighbourhood (Fig. 4b), a smoothing technique is needed to reconstruct the cells shape.



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

Fig. 4. Example of: (a) cell, (b) hue difference to background and (c) its smoothed out version with (d) localized *highest* points (*).

The smoothing operation in our approach relies on the fact that this sort of 2D signal can be modelled as a sum of sinusoids [4] with defined amplitudes, phase shifts and frequencies. Cutting low amplitude frequencies off (leaving only a few significant ones with highest amplitude) will result in a signal deprived of our problematic local noise effect (Fig. 4c).

Localization of objects on a modelled map of cells can be performed locally using evolutionary (1+1) search strategy [1]. A population of individuals is thus searching for a maximum in the map (Fig. 4c) starting from points defined by pre-segmentation result. The change for the better position of an individual is calculated as a product of randomly generated distance with normal distribution $N(0,1)$ and an exponentially decreasing in time radius $R_t = R_{\max}(1/R_{\max})^{t/t_{\max}}$, where $R_{\max}$ is the maximal radius detected by the Hough transform. Performed experiments shown that just several epochs are needed to localize all the cells, what requires small computational effort.

3.2. THE ACTIVE CONTOURING METHOD

The Hough transform can result in either undersegmented images or missed some nucleuses. However it can be observed that pre-segmentation with low *HT* threshold values finds all areas containing nucleuses. Boundary of the area extracted with the *HT* is then used as an initial contour shape for further segmentation using active contour method. Shrinking contour will have to split into multiple contours to separate multiple objects. Such behaviour is very hard to achieve using classical marker-based active contour. The solution to this problem is application of Fast Marching Method (FMM), developed by Sethian [8], which handles very well changes in topology of the contour. The problem with the original FMM is that the contour can be moved only in one direction. This means that any error in segmentation cannot be corrected and algorithm requires additional stop condition. To deal with this problem, multilabel extension to the classical FMM [9,10] was proposed.

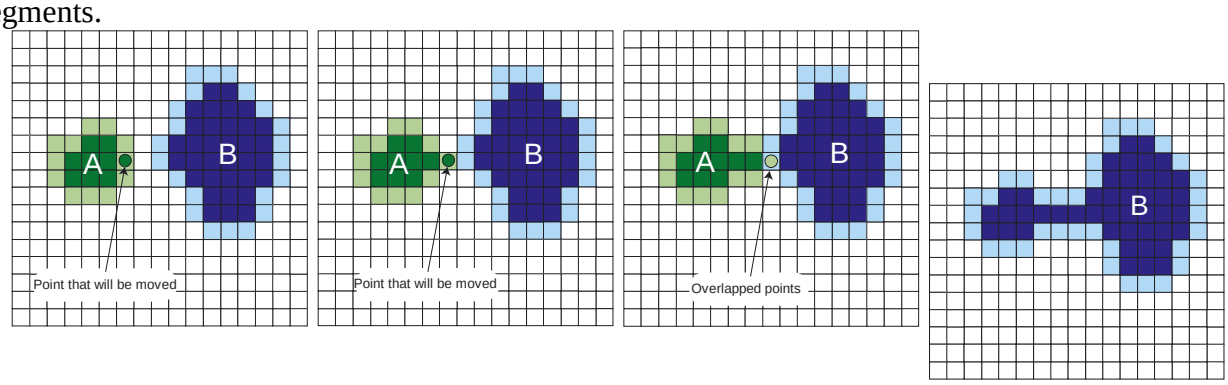
The multilabel FMM is initialized by mask composed with pre-segmentation result from the Hough transform and detected interiors of the nucleuses. Each connected component of the mask is marked with different label and its border will be an initial state for propagating contours. What is important, initialization mask does not have to be perfect. It means that one nucleus can contain several initialization areas and false centres in the background area are acceptable.

Initial contour propagation is similar to original FMM method. Expansion of the contour is governed by a propagation speed defined globally for all the contours. Speed is based on the difference between mean colour in the initialization area and colour of the pixel under the contour:

$$F = \frac{1}{\left|g(x,y) - \bar{g}(i)\right|^3 + 1}, \tag{2}$$

where $g(x,y)$ is the colour under the contour and $\bar{g}(i)$ is the mean colour under the i-th segment. Such a speed definition slows down the contour near the detected object boundary what increases probability of contours meeting near nucleus boundary.

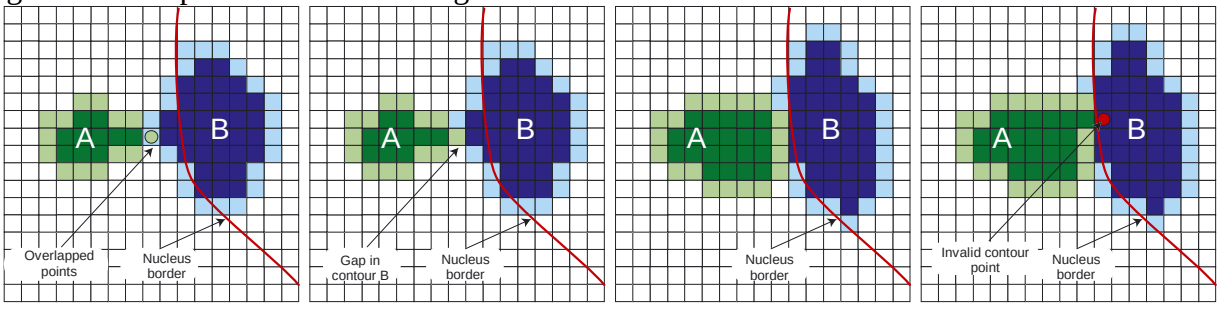
When two segments meet, mean colour of the segments is compared. Comparison is taken at the point where contours start to overlap. When difference between mean colours from these two segments is below certain threshold segments are merged into one (Fig. 5). To ensure maximum efficiency, labels from the smaller segment are changed to the value of those from the larger segment. Additionally, new mean colour for the segment is calculated from mean colours of connected segments.



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

Fig. 5. Segments merging: (a) segments one pixel away, (b) connected segments, (c) overlapped point, (d) connected segments.

If two segments that meet are not classified to be merged, the propagating segment can push back another segment under certain circumstances. At the meeting point differences between current pixel colour and mean colour of each segment is compared. Segment with lower difference value wins and replaces current label with its own. Replacement is performed as long as condition is meet. Contour that was pushed back cannot be propagated farther at places where its labels was replaced by another contour (Fig. 6). Contour points that cannot be moved are no longer considered during calculations. Since contour can be pushed back only once, there is no oscillation at the object boundary known from the classical active contour methods. Additionally reduction of the contour length increases performance of the algorithm.

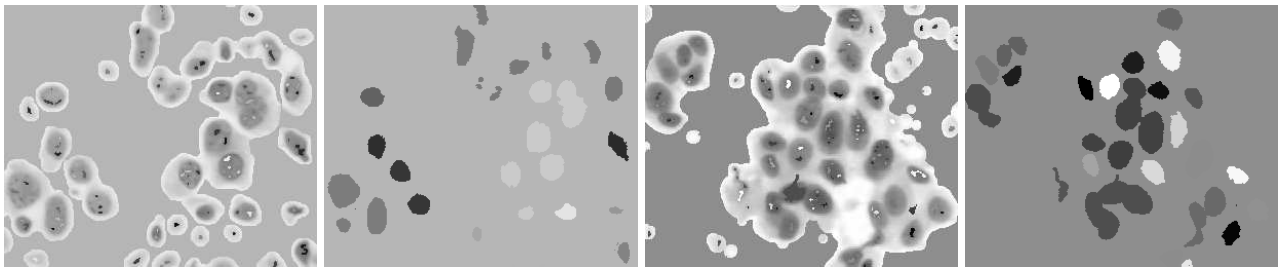


(a) (b) (c) (d)

Fig. 6. Segments pushing: (a) differences at the meeting point are calculated, (b) contour point is replaced, (c) segment A approaches nucleus boundary, (d) segment A cannot move farther.

3.3. SUMMARY AND EXEMPLARY RESULTS

Algorithm requires properly exposed pictures of microscopic samples. However, it is stable enough to perform well for slightly under- or overexposed images. Additionally, initialization step does not have to be very precise. Initialization errors such as multiple seeds inside one nucleus or seeds at the nucleus border does not influence segmentation quality. Figure 7 shows exemplary results of segmentation. As can be seen, even small nucleuses are detected while other parts of cells are ignored. Only nucleuses that are visually connected on the image are detected as one segment. Segmentation of a 704 x 567 pixels image takes several seconds on Athlon 1.4 GHz processor.



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

Fig. 7. (a) and (c) – initialization overlaid on segmented image, (b) and (d) – results of the segmentation.

3. CONCLUSIONS

Conducted experiments shown that the Hough transform and the evolutionary algorithms can be effectively used for initialization of active contour method. Modification of the multilabel FMM used for experiments turned out to be very stable and robust to initialization errors. Visually assessed segmentation quality is promising and gives good detection of even small objects. Additionally, the shape of nucleuses was represented accurately. There are still some problems requiring further research. One of them is proper selection of merging threshold and another one is detection of overlapping nucleuses. Future research includes larger test set and quality assessment based on hand-prepared reference segmentation. Presented technique is meant for initial processing for automated cytological diagnosis system.

BIBLIOGRAPHY

[1] ARABAS J., Lectures on Evolutionary Algorithms, WNT, 2004 (in Polish).
[2] GONZALEZ R., WOODS R., Digital Image Processing, Prentice Hall, 2002.
[3] KIMMEL M., LACHOWICZ A., ŚWIERNIAK A. (Eds.), Cancer Growth and Progression, Mathematical Problems and Computer Simulations, Int. J. of Appl. Math. and Comput. Sci., Vol. 13, No. 3, Special Issue, 2003.
[4] MADISETTI V., WILLIAMS D., The Digital Signal Processing Handbook, CRC Press, 1997.
[5] MARCINIAK A., OBUCHOWICZ A., MONCZAK R., KOŁODZIŃSKI M., Cytomorphometry of Fine Needle Biopsy Material from the Breast Cancer, Proc. of the 4th Int. Conf. on Comp. Recogn. Systems CORES'05, Adv. in Soft Computing, pp. 603-609, Springer, 2005.
[6] PRATT W., Digital Image Processing, John Wiley & Sons, 2001.

[7] RUSS J., The Image Processing Handbook, CRC Press, 1999.
[8] SETHIAN J., Fast marching methods, SIAM Review, Vol. 41, No. 2, 1999.
[9] STEĆ P., DOMAŃSKI M., Video Frame Segmentation Using Competitive Contours, Proc. of the 13th European Signal Processing Conference, Antalya, Turkey, 2005.
[10] STEĆ P., Segmentation of Colour Video Sequences Using the Fast Marching Method, University of Zielona Góra Press, Poland, 2005.
[11] TADEUSIEWICZ R., Vision Systems of Industrial Robots, WNT, 1992 (in Polish).
[12] TOFT P., The Radon Transform, Ph.D. Thesis, Technical University of Denmark, 1996.
[13] ŻORSKI W., Image Segmentation Methods Based on the Hough Transform, Studio GiZ W-wa, 2000 (in Polish).