Maciej HREBIEŃ[*], Piotr STEĆ[*]

FINE NEEDLE BIOPSY MATERIAL SEGMENTATION WITH HOUGH TRANSFORM AND ACTIVE CONTOURING TECHNIQUE

This paper describes a hybrid segmentation method of cytological images for their analysis by means of Hough transform and Active Contours methods. One also can find there a short description of image pre-processing that was recommended for preliminary experiments organisation. The results were collected by a specifically prepared benchmarking database.

1. INTRODUCTION

Automatic diagnostic systems of cancer diseases have a very long history and are still under development by many research and university centres as well as various commercial institutions [3]. Since a breast cancer is becoming the most common women's disease today, many efforts are still undertaken; to increase its early detecting.

The nucleus of the cell is the place where a breast cancer malignancy can be observed. Therefore, it is crucial problem for camera-based automatic diagnostic system; separating the cancer 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 processing and real-time operations implementation, the highest diagnostic level is expected. The methods with deprived drawbacks of already known approaches have to be recognised.

The paper discusses a hybrid method that is combining the Hough transformation with the Active Contour method, for cytological image segmentation. The paper also presents some pre-processing steps used in the approach for preliminary experimental procedures collecting some sampling data by the elaborated benchmark database.

[*]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

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 be removed to reduce processing complexity. An RGB colour image can be converted to a 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 has a low contrast, an enhancement technique is needed to improve their quality. In the presented investigations a simple histogram processing was used; with linear transformation of images intensities level, namely the 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 shapes. 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 very complex. The shape of ellipses can be approximated by a given number of circles. Anyhow, detection of circles is simpler, in a sense of required computing steps. They use only one parameter in the computing algorithm; radius R.

The Hough transform [11, 12] can be easily adopted for the purpose of circles detection. This transformation in the discrete space could be defined by:

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

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 a level of similarity of image g to circle placed at the $(\hat{i},\hat{j})$ position and defined by the radius R. The accumulator is thus three dimensional.

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 [10] or theirs heavy or light versions (Fig. 1).

An example result of the gradient estimation, using Sobel's mask and an example *HT* accumulator intersection for $R = 9$ is given in Fig. 2. One can easy observe the accumulator

containing visible higher values in places where the feature image *g* is similar to a circle. The warm coloured peaks correspond to circle centres for a given radius *R*.

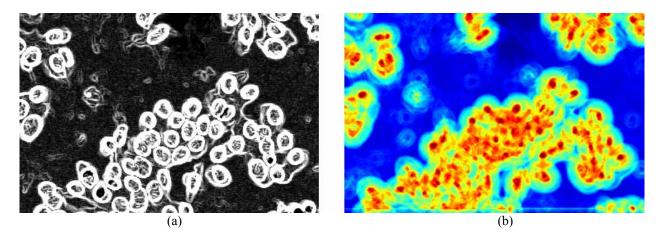(a)   (b)

Fig. 2. An exemplary gradient estimation (a) and *HT* accumulator intersection (b) for circles with the radius *R* = 9 (warmer colours correspond to higher similarity of a given region to circle).

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. 3).

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 removed. Additionally, the method can be easily decomposed to perform even more effectively on SIMD organized machines, parallelised for multithreaded systems and even implemented in dedicated hardware because of its simple complexity.
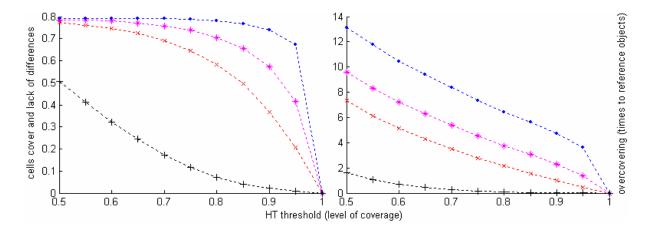

Fig. 3. Charts illustrating: influence of HT threshold value on cell cover and lack of differences (left) and overcovering (right) for Prewitt (×), Sobel (*), heavy (•) and light (+) base gradient masks (experiments performed on a randomly selected 346 element Zielona Góra's ONKOMED [5] cytological benchmark database for radiuses in the 4-21 range).

The example results of image pre-segmentation, for different HT threshold values, are shown in Fig. 4. As one can easily notice lower threshold values give us a very good approximation of regions containing objects of interest – namely cells.
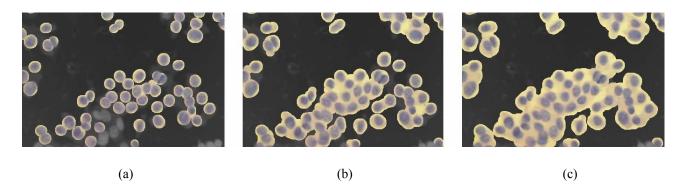

(a)   (b)   (c)

Fig. 4. Exemplary results of image pre-segmentation for: (a) high, (b) middle and (c) low *HT* threshold values.

3. IMAGE SEGMENTATION

3.1. CELLS LOCALISATION

The results obtained from the pre-segmentation mechanism can lead us to the estimation of average background colour. This information can be used for cells modelling as a colour distance between background and objects what fulfils requirements of lack of any colour dependency in image object (the colour of contrasting pigment may be changed in future). In the discussed approach the HSV colour space was used where the distance was defined, as the absolute Hue value difference between background and objects.

Because the distance can vary in local neighbourhood (Fig. 5b), a smoothing technique is needed for reconstructing the cells' shape.

(a)   (b)   (c)   (d)
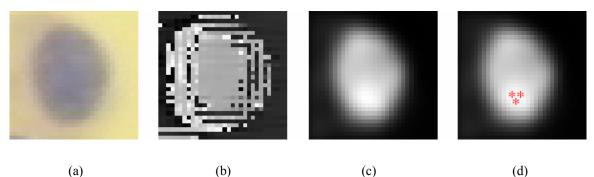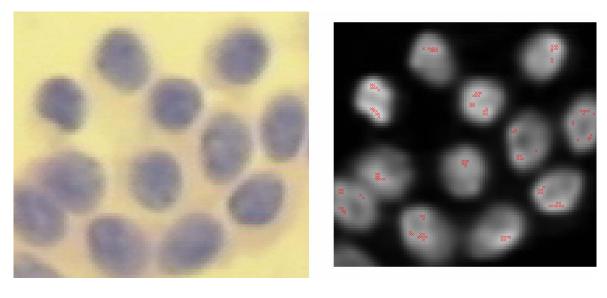
Fig. 5. Example of: (a) cell, (b) hue difference to background and (c) its smoothed out
version with (d) localized *highest* points (*).

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 strictly defined amplitudes, phase shifts and frequencies. Cutting the low amplitude frequencies off (leaving only a few significant signals, with the highest amplitude) provides the result in a signal deprived of our problematic local noise effect (Fig. 5c).

Localisation 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. 5c), starting from points defined by pre-segmentation result.

The better position finding, of an individual, is done by calculating a product of randomly generated distance, with a 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.

The performed experiments shown that several epochs are needed to localise all the cells (see for instance Fig. 6), what requires simpler computational efforts. Additionally, the influence of a spot illumination in the image, are results modelled by a map being *higher* in centre and *lower* on the corners, eliminated from this object.

Fig. 6. A fragment of original image (a) and an exemplary result (b) of cell localization points (red asterisks).

3.2. THE ACTIVE CONTOURING METHOD

The Hough transformation can result in either under-segmented images or with 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.

This behaviour is very hard to achieve using classical marker-based active contour. The solution to this problem was identified by application of Fast Marching Method (FMM), developed by Sethian [8], which handles with a very well changed contour in topology. 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 break condition. To deal with this problem, multi-label extension, to the classical FMM [9], was considered.

The multi-label FMM is initialised by mask composed with pre-segmentation result from the Hough transformation 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, initialisation mask does not have to be perfect. It means that one nucleus can contain several initialisation areas and false centres in the background area are acceptable.

Initial contour propagation is similar to original FMM method. All segments initialised earlier are propagated outwards using a modified fast marching algorithm. The segment labels for the points visited by contours are positive integers. The points of each contour are marked with negative numbers of segment labels. All contour points from all segments are included into the same sorted list. Thanks to this, no additional time synchronisation between the segments is required.
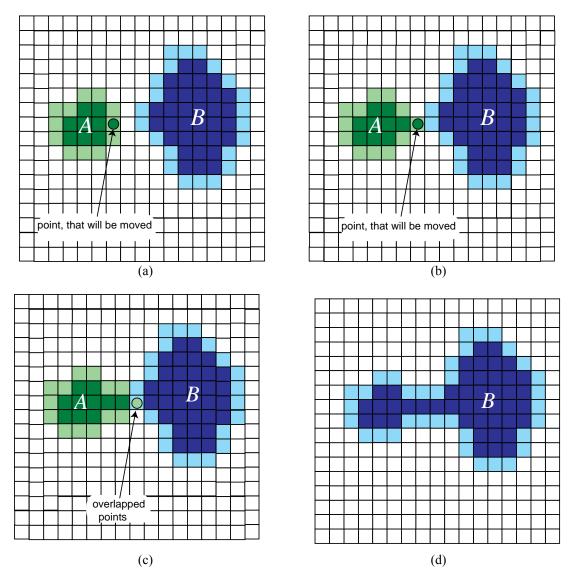

Fig. 7. Segments merging: (a) segments one pixel away, (b) connected segments,
(c) overlapped point, (d) connected segments.

This situation is naturally handled by the fast marching algorithm since it can propagate contours of any topology. At this stage of propagation, there is in fact no difference between the standard and the multi-label implementation apart from the fact that

the new label for the contour point is inherited from the segment that propagates at the current algorithm step.

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 initialisation area and colour of the pixel under the contour:

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

(2)

where $g(x,y)$ is the colour under the contour and $\overline{g}(i)$ 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.
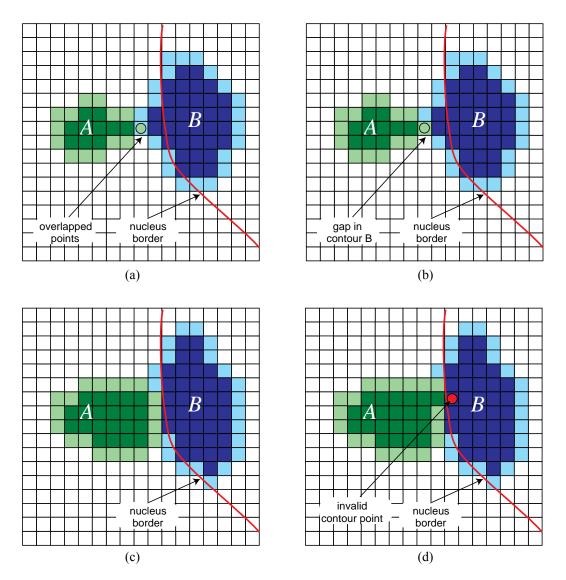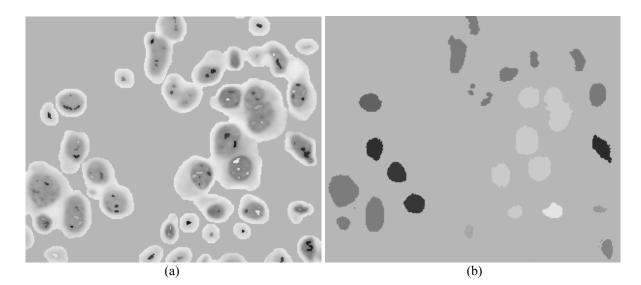

Fig. 8. Segments pushing: (a) differences at the meeting point are calculated, (b) contour point is replaced, (c) segment A approaches nucleus boundary, (d) segment A cannot move farther.

31

When two segments meet, mean colour of the segments is compared. Comparison is taken at the point where contours start to overlap. When these differences, between mean colours of these two segments, are below a certain threshold segments, they are merged into one (Fig. 7). To ensure maximum efficiency, labels from the smaller segment are changed to the value of those from the larger segment. Also, contour points from smaller segments are assigned with the value from larger segments. Additionally, new mean colour for the segment is calculated from mean colours of 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 met. Contour that was pushed back cannot be propagated farther at places where its labels were replaced by another contour (Fig. 8).
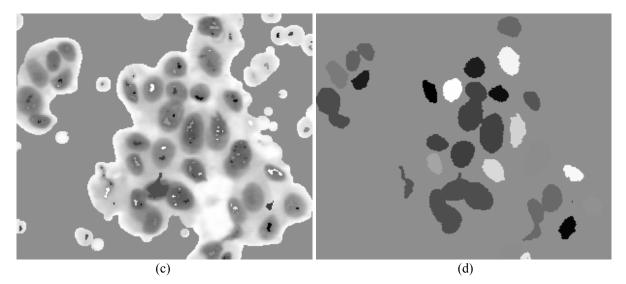

(a)

(b)

(c)

(d)

Fig. 9. (a) and (c) – initialisation overlaid on segmented image, (b) and (d) – results of the segmentation.

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, the reduction of the contour length increases performance of the algorithm.

When a contour has no possibility of propagating further, no new points are added to the contour. This implies the reduction of the total length of the sorted list used by the fast marching algorithm and the same performance improvement. The presented algorithm stops propagation when all image points are assigned to segments and there is no segment that could push back another segment. The algorithm cannot run infinitely because oscillations between segments are impossible. No segment can visit twice the same area. Namely, when a segment was pushed back by another segment, it cannot get back the lost pixels.

3.3. SUMMARY AND SOME 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. Initialisation errors such as multiple seeds inside one nucleus or seeds at the nucleus border do not influence segmentation quality. Fig. 9 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. The final segmentation of a 704 x 567 pixels pre-processed image takes several seconds on Athlon 1.4 GHz processor (about 15-30 seconds depending on image content).

4. CONCLUSIONS

The conducted experiments shown that the Hough transformation and the evolutionary algorithms can be effectively used for initialisation of active contour method. Modification of the multi-label FMM used for experiments turned out to be very stable and robust to initialisation 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 investigations. One of them concerns a proper selection of merging threshold and another one concerns detection of overlapping nucleuses.

Future works will include larger test-set with quality assessment-based on hand-prepared reference segmentations.

The presented technique is a meant for initial processing for the automated cytological diagnostic 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] TADEUSIEWICZ R., Vision Systems of Industrial Robots, WNT, 1992 (in Polish).

[11] TOFT P., The Radon Transform, Ph.D. Thesis, Technical University of Denmark, 1996.

[12] ŻORSKI W., Image Segmentation Methods Based on the Hough Transform, Studio GiZ W-wa, 2000 (in Polish).