Automatic nuclei detection on cytological images using the firefly optimization algorithm

Paweł Filipczuk, Weronika Wojtak and Andrzej Obuchowicz

Abstract The firefly algorithm is a powerful optimization method inspired by the flashing behavior of fireflies. In our work on computer aided breast cancer diagnosis we met a problem of automatic marking of nuclei. Our system is based on analysis of microscopic images of fine needle biopsy material. The task of the system is to identify benign and malignant lesions (optionally it can also distinguish fibroadenoma). For this purpose it extracts nuclei from cytological images in segmentation phase, then it determines their morphometric features and finally classifies the case. Some segmentation methods require a preliminary selection of objects on the image. We have adapted the firefly algorithm to this task. We have also proposed an initialization procedure. The method was experimentally shown to be satisfactorily effective. The approach was tested with real case medical data collected from patients of the Regional Hospital in Zielona Gora.

1 Introduction

According to the National Cancer Registry in Poland breast cancer is the most common cancer among women. In 2009, there were 15,752 diagnosed cases of breast cancer in Polish women. Out of these cases, 5242 deaths were the result. There has also been an increase of breast cancer by 3-4% a year since the 1980's. The effec-

Paweł Filipczuk
Institute of Control & Computation Engineering, ul. Podgórna 50, 65-246 Zielona Góra, Poland e-mail: p.filipczuk@issi.uz.zgora.pl

Weronika Wojtak
Faculty of Electrical Engineering, Computer Science and Telecommunications, ul. Podgórna 50, 65-246 Zielona Góra, Poland e-mail: w.wojtak@gmail.com

Andrzej Obuchowicz
Institute of Control & Computation Engineering, ul. Podgórna 50, 65-246 Zielona Góra, Poland e-mail: a.obuchowicz@issi.uz.zgora.pl

tiveness of treatment largely depends on early detection of the cancer. Important and often used diagnostic method is so-called triple-test. It is based on three medical examinations and allows to achieve high confidence of diagnosis. The triple-test includes self examination (palpation), mammography or ultrasonography imaging and Fine Needle Biopsy (FNB) [12]. FNB is an examination consisting in obtaining cytological material directly from tumor. Collected material is examined under a microscope to determine the prevalence of cancer cells. The present approach requires a deep knowledge and experience of the cytologist responsible for the diagnosis. Automatic morphometric diagnosis can make the decision objective and assist unexperienced specialist. It can also allow screening on a large scale where only difficult and uncertain cases would require additional human attention. Along with the development of advanced vision systems and computer science, quantitative cytopathology has become a useful method for the detection of diseases, infections as well as many other disorders [5, 11]. In the literature one can also find approaches to the breast cancer classification [4, 1, 6, 7, 9, 14].

The paper presents a part of work in progress on a system of automatic breast cancer diagnostic system based on analysis of cytological images of FNB material. The system includes 3 steps: preprocessing, segmentation and classification. As a result it marks a case as benign or malignant (additionally it might also distinguish fibroadenoma). This article is focused on segmentation phase. Some segmentation algorithms such as the marker-controlled watershed or GrowCut require to be initialized by indicating objects of interest on the image [13, 15]. Moreover, using our previously developed segmentation methods based on clustering we often encountered joined nuclei [2, 3, 8]. Additional information about quantity and location of nuclei is very useful in task of separation individual objects. In this paper we present method for automatic nuclei detection. The approach is based on the firefly optimization algorithm and was tested on real case medical data with promising results.

The paper is divided into four sections. Sect. 1 gives an overview of breast cancer diagnosis techniques. Sect. 2 describes the process of acquisition of medical images used to test the system. Sect. 3 presents in detail the algorithm detecting nuclei on images. Sect. 4 shows the experimental results obtained using the proposed approach. The last part of the work includes conclusions and bibliography.

2 Medical Images Database

The database contains 750 images of the cytological material obtained by FNB. The material was collected from 75 patients of the Regional Hospital in Zielona Gora. It gives 10 images per case which was recommended amount by the specialists from the hospital [1, 10] and allows correct diagnosis by a pathologist. The set contains 25 benign, 25 malignant and 25 fibroadenoma cases. Biopsy without aspiration was performed under the control of ultrasonograph with a 0.5 mm diameter needle. Smears from the material were fixed in spray fixative (Cellfix of Shandon

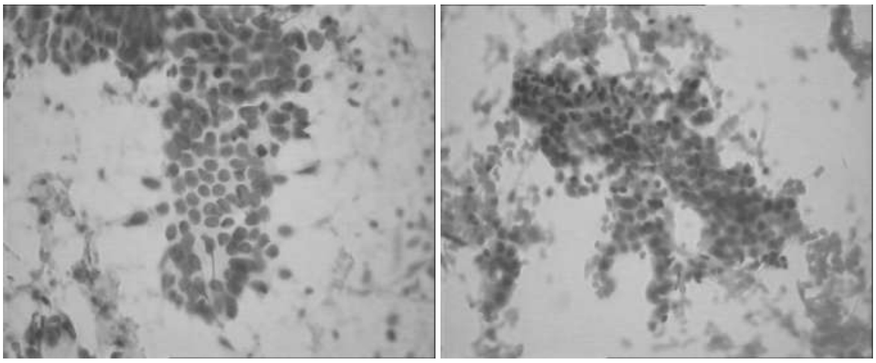

Fig. 1 Sample microscopic images of breast tumor

company) and dyed with hematoxylin and eosin (h+e). The time between preparation of smears and their preserving in fixative never exceeded three seconds. The images were recorded by SONY CDD IRIS color video camera mounted atop AXIOPHOT microscope. The slides were projected into the camera with 40x and 160x objective and a 2.5x ocular. Each case consists 1 overall image generated for enlargement 100x and 9 for enlargement 400x. Images are BMP files, 704x578 pixels, 8 bit/channel RGB. All cancers were histologically confirmed and all patients with benign disease were either biopsied or followed for a year. Sample images are presented on Fig. 1.

3 Automatic nuclei detection

The aim of presented approach is to assign a marker to each nuclei on the image. We decided to adapt the firefly algorithm for this task. The algorithm is an optimization technique inspired by the flashing behavior of fireflies [16]. The idea lays to use a swarm of simple agents which coordinate their activities to solve a complex problem. Since nuclei are the darkest objects on the image our problem boils down to find local minimums of luminance in the image space. At first we initialize the algorithm by generating initial population of fireflies. Then the members of the population are iteratively moved according to the principle set forth below.

3.1 Initialization

The algorithm needs to be initialized by starting distribution of fireflies. The initial population should give the best possible arrangement of cells in the image. For this purpose we have proposed following procedure. At first the nuclei are extracted from

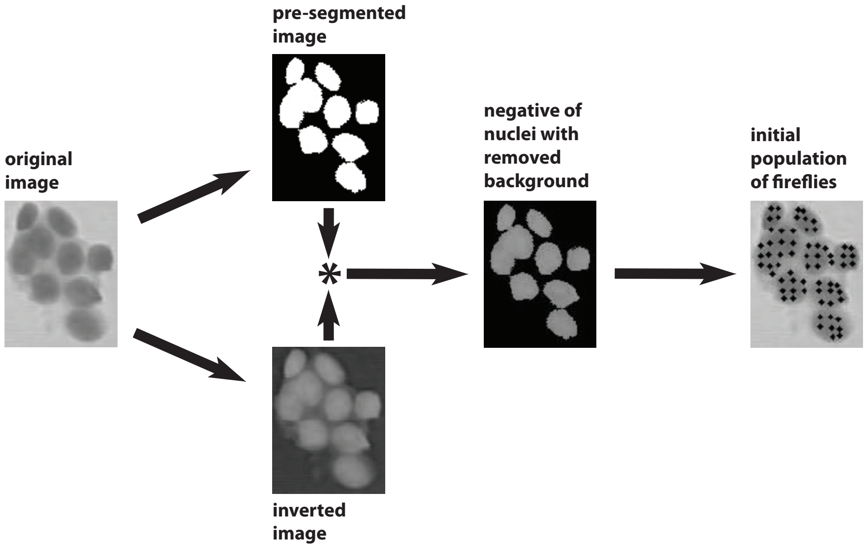

Fig. 2 The initialization procedure diagram

the background in pre-segmentation step. We have tested 8 methods based mainly on thresholding, clusterization and combination of both: AT (Adaptive Thresholding), KM (K-Means), FCM (Fuzzy C-Means), CLN (Competitive Learning Networks), ATmKM, ATmFCM and ATmCLN (last three are combination of Adaptive Thresholding used to remove the background and K-Means, Fuzzy C-Means or Competitive Learning Networks respectively to distinguish nuclei from red blood cells). Above-mentioned methods are described in our previous articles. Further details one might find in following references [2, 3, 8].

Each firefly is represented by its position and lightness. Since the nuclei are the darkest objects on the image we assume that lightness of given firefly is a negation of values of pixels represented by the firefly (the darker pixels the brighter firefly). The negative of the image is multiplied by the results of pre-segmentation. That operation removes the background. The next step is to divide the image into square blocks of a certain size. For each block containing at least one nonzero pixel (the background is removed and its value is 0) a single firefly is determined. The position of the firefly is centroid computed for the block. Lightness is the mean value of pixels. The initialization procedure diagram is presented on Fig. 2.

3.2 The algorithm

Each firefly interacts with each other with some strength. The attractiveness is proportional to the light intensity, so each firefly is attracted by its neighbor that glows brighter. The light intensity $I(r)$ varies according to the inverse square law:

$$I(r) = \frac{I_s}{r^2} , \tag{1}$$

where I_s is the intensity at the source and r is the distance between two fireflies. Light is absorbed in the media with an absorbtion coefficient γ. The combination of the inverse square law and the absorbtion can be approximated using to avoid singularity at $r = 0$ in I_s/r^2:

$$I(r) = I_0 e^{-\gamma r} , \tag{2}$$

where I_0 is the original light intensity. As the firefly attractiveness is proportional to the brightness seen by its neighbor, the attractiveness function is determined by

$$\beta(r) = \beta_0 e^{-\gamma r} , \tag{3}$$

where β_0 is the attractiveness at $r = 0$. To determine the distance affecting the attractiveness between any two fireflies i and j at positions x_i and x_j Euclidean measure is used:

$$r_{ij} = \|x_i - x_j\| . \tag{4}$$

In each iteration less attractive fireflies move to the brighter ones. The movement of the firefly i at location x_i to firefly j at location x_j is expressed by

$$x_i(t+1) = x_i(t) + \beta_0 e^{-\gamma r}(x_j - x_i) . \tag{5}$$

Iterations are performed until reaching maximum number of iterations or meeting the stop condition. In each iteration maximum movement is determined and if the value is smaller than given epsilon, the procedure stops. The result are groups of fireflies gathered in centers of nuclei. Then fireflies that lay close to each other are merged to achieve final markers.

4 Experimental results

The algorithm was tested by comparison with results obtained manually. There was prepared a set of 351 manually segmented images, of which 145 were benign and 206 were malignant. For each image the algorithm was performed and achieved markers were assigned to one of three classes. The first class consisted markers pointing out the background. The second were points correctly indicating nuclei (points which alone indicate a particular nucleus). The last one were points indicating nuclei, but there was at least one another point indicating the same nucleus. The final rate was a percentage of markers correctly indicating nuclei to the total number of markers.

To find optimal parameters values there were three sequences of tests performed. The aim of the first one was to determine the best pre-segmentation method. There were 8 methods described in Sect. 3. The second sequence was conducted to find optimal block size. Finally, the algorithm was performed for several γ values. The

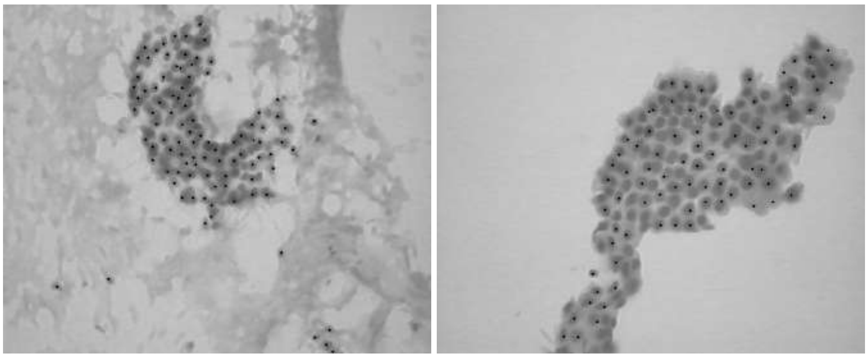

Fig. 3 Example results - original image with determined markers

Table 1 Efficiency for different methods of pre-segmentation (abbreviations description in Sect. 4), where block size is 6 and γ is 0.03

AT	KM	FCM	CLN	ATmKM	ATmFCM	ATmCLN
26.05%	65.32%	67.56%	54.70%	67.62%	69.63%	57.05%

Table 2 Efficiency for different block sizes, where pre-segmentation method is ATmFCM and γ is 0.03

3	6	9	12	15
68.97%	69.63%	67.85%	65.60%	59.62

Table 3 Efficiency for different γ values, where pre-segmentation method is ATmFCM and block size is 6

0.0013	0.0025	0.005	0.01	0.015	0.02
69.24%	71.48%	72.72%	73.55%	74.44%	72.00%
0.025	0.03	0.035	0.04	0.045	0.05
70.61%	69.63%	65.04%	61.69%	56.30%	53.05%

average results are presented in Tables 1, 2 and 3. Sample images with determined markers are presented in Fig. 3.

The final average rate was 74.44% of correct markers, 23.86% of markers indicating background and 1.7% of markers indicating nucleus, which was pointed out by at least one another marker. However, it must be mentioned that the number of points indicating background is inflated due to the fact that during manual segmentation blurred and densely clustered nuclei were omitted. The discrepancy has been manually estimated on 10% randomly selected images and after correction the average rate was 82.2%.

5 Conclusions

The aim of the research was to propose a method for automatic nuclei detection on cytological images of FNB. For this purpose we have adapted the firefly optimization algorithm. We have also developed an initialization procedure. The approach was experimentally shown to be satisfactorily effective. Despite the fact that the experimental results were understated due to the testing procedure the average percentage of correct markers to the total number of all markers for 351 images was over 74%.

The method was described in context of the breast cancer diagnostic system we are working on. However, we suppose its applicability might be much broader. In the future research we are going to apply the approach to initialize several segmentation methods requiring a preliminary selection of objects on the image.

Acknowledgement

The author is a scholar within Sub-measure 8.2.2 Regional Innovation Strategies, Measure 8.2 Transfer of knowledge, Priority VIII Regional human resources for the economy Human Capital Operational Programme co-financed by European Social Fund and state budget.

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Lubuskie
Worth your while
EUROPEAN UNION
EUROPEAN
SOCIAL FUND

References

1. P. Filipczuk, M. Kowal, and A. Marciniak. Feature selection for breast cancer malignancy classification problem. *J. Medical Informatics & Technologies*, 15:193–199, 2010.
2. P. Filipczuk, M. Kowal, and A. Obuchowicz. Automatic breast cancer diagnostis based on k-means clustering and adaptive thresholding hybrid segmentation. In ed. Ryszard S. Choraś, editor, *Image processing and communications challenges 3, ISBN: 978-3-642-23153-7*, Advances in Intelligent and Soft Computing : 102, pages 295–303. Springer-Verlag, Berlin - Heidelberg, 2011.
3. P. Filipczuk, M. Kowal, and A. Obuchowicz. Fuzzy clustering and adaptive thresholding based segmentation method for breast cancer diagnosis. In M. Woźniak A. Żołnierek red. R. Burduk, M. Kurzyński, editor, *Computer recognition systems 4, ISBN: 978-3-642-20319-0*, Advances in Intelligent and Soft Computing : 95, pages 613–622. Springer-Verlag, Berlin - Heidelberg, 2011.
4. J. Gil, H. Wu, and B. Y. Wang. Image analysis and morphometry in the diagnosis. *J. Microsc. Res. Tech.*, 59:109–118, 2002.

5. M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147–171, 2009.
6. M. Hrebień, P. Steć, A. Obuchowicz, and T. Nieczkowski. Segmentation of breast cancer fine needle biopsy cytological images. *Int. J. Appl. Math and Comp. Sci.*, 18(2):159–170, 2010.
7. L. Jeleń, T. Fevens, and A. Krzyżak. Classification of breast cancer malignancy using cytological images of fine needle aspiration biopsies. *Int. J. Appl. Math and Comp. Sci.*, 18(1):75–83, 2010.
8. M. Kowal, P. Filipczuk, and J. Korbicz. Hybrid cytological image segmentation method based on competitive neural network and adaptive thresholding. *Pomiary, Automatyka, Kontrola*, Vol. 57(nr 11):1448–1451, 2011.
9. M. Kowal and J. Korbicz. Segmentation of breast cancer fine needle biopsy cytological images using fuzzy clustering. In J. Kornacki, Z. Raś, S. Wierzchoń, and J. Kacprzyk, editors, *Advances in Machine Learning I*, pages 405–417. Springer-Verlag, Berlin-Heidelberg, 2010.
10. A. Marciniak, A. Obuchowicz, A. Monczak, and M. Kołodziński. Cytomorphometry of fine needle biopsy material from the breast cancer. In *Proc. 4th Int. Conf. on Computer Recognition Systems CORES' 05*, pages 603–609, 2005.
11. J. Śmietanski, R. Tadeusiewicz, and E. Łuczyńska. Texture analysis in perfusion images of prostate cancer - a case study. *Int. J. Appl. Math and Comp. Sci.*, 20(1):149—-156, 2010.
12. J. C. E. Underwood. *Introduction to biopsy interpretation and surgical pathology*. Springer-Verlag, London, 1987.
13. Automata Vladimir Vezhnevets. Growcut" - interactive multi-label n-d image segmentation by cellular, 2005.
14. W. H. Wolberg, W. N. Street, and O. L. Mangasarian. Breast cytology diagnosis via digital image analysis. *Analytical and Quantitative Cytology and Histology*, 15:396–404, 1993.
15. Houqiang Li Xiaodong Yang and Xiaobo Zhou;. Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy. *IEEE Trans. on Circuits and Systems I: Regular Papers*, 53(11):2405–2414, 2006.
16. X.S. Yang. *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, Frome, 2010.