An ensemble algorithm for the nuclei segmentation in the histological images

6 authors, including:

Piotr Krajewski

Dolnoslaskie Centrum Analizy Danych

2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Witold Dyrka

Wroclaw University of Science and Technol…

16 PUBLICATIONS **64** CITATIONS

SEE PROFILE

An ensemble algorithm for the nuclei segmentation in the histological images

Jakub Czakon[*1], Piotr Giedziun[1], Michał Błach[2], Grzegorz Żurek[1,3],
Piotr Krajewski[†1], and Witold Dyrka[‡1,2]

[1]Stermedia Sp. z o. o. Wroclaw, Poland
[2]Department of Biomedical Engineering, Wroclaw University of Technology, Poland
[3]Faculty of Mathematics, Wroclaw University of Technology, Poland

September 28, 2015

1   Introduction

Image analysis for computer-aided medical diagnosis is a rapidly growing discipline of scientific research. Many medical tasks, like segmenting MRI images of the brain or counting red and white blood cells, can be conducted in an automated fashion, which not only is financially beneficial but above all permits doctors to devote more time to patients. In the era of cloud computing and deep learning more and more problems that were previously regarded as too complex for computers to handle are being solved and tools that implement those solutions are being delivered to the society. However, there are a variety of challenges, especially in contemporary medicine, that have not been properly worked out yet. Without a doubt, many of those challenges can be found in the field of digital pathology. In this paper we focus specifically on the problem of segmentation of histological images. Image segmentation is one of the most generic problems in the field of visual content processing. Generally it is defined as assigning each pixel of the n-by-m image matrix an element from the k-element set of labels. In the context of this project the task consists of assigning each pixel to a specific cluster, or a group of spatially connected pixels that share similar color or intensity.

2   Methodology

2.1   Preprocessing

The training dataset consisted of 14 unsegmented histological images at which nuclei were annotated by an expert pathologist. Images were taken at two different

---

[*]j.czakon@stermedia.pl
[†]p.krajewski@stermedia.pl
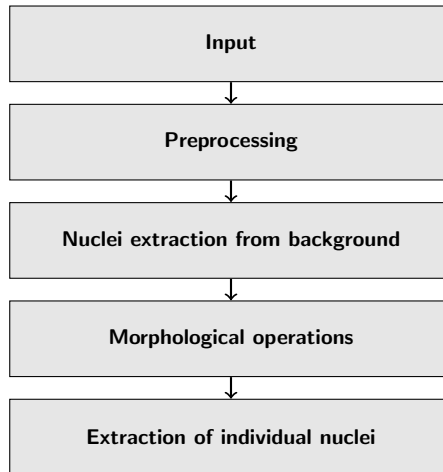[‡]witold.dyrka@pwr.edu.pl

Figure 1: Overview of the algorithm.

magnifications and ranged in size from 459 by 392 pixels to 1032 by 808 pixels. Importantly, they varied substantially in terms of coloring, intensity and in number, size and density of nuclei. To deal with this diversity, we utilized a two-stage image preprocessing. First, we applied Reinhard's normalization transform [1] to original RGB images in order to adjust color histograms to that of a selected target image. Then, we extracted the Hematoxylin channel [2] from the transformed RGB image.

2.2 Nuclei extraction from background

The nuclei extraction procedure was divided in two steps: (1) extracting Regions of Interest (RoI) - containing putative, potentially glued nuclei - from background; (2) decomposing RoIs into separate nuclei. In the first step, we opted for combining predictions obtained from multiple foreground extraction algorithms. We used two methods relying only on color intensities: the classical K-Means (KM) [3] and Gaussian Mixture Model (GMM) [4]. Moreover, we implemented two algorithms aware of spatial information: Spatial Distance Weighted Fuzzy C-Means (SDWFCM) [5] and Dictionary Model (DM) [6]. In our final solutions, we used predictions obtained using three or four (only DM) different parametrizations for each model. The parametrizations were chosen by fitting models to the training dataset and optimizing the DICE1 score.

2.3 Morphological operations

In the next step, the binary masks were subjected to morphological operations in the manner inspired by the work of Veta et al. [7]. It consisted of smoothing RoIs along with removal of small objects by performing the combination of two morphological closing operations with different sizes of structuring elements (SE), one with a large SE to remove small objects and the other with small SE to clean up the resulting clusters. Moreover, since choosing universal SE sizes for all images is practically impossible, we generated an ensemble of morphological closing outputs with 5 different sets of SE sizes. Such approach has been recently shown to improve the results of image segmentation [8].

Finally, binary masks - generated using different segmentation methods and transformed using various sets of SE - were combined in order to generate the single most probable binary labeling for each image. Specifically, the nuclei probability maps for each image were calculated by overlapping all masks, and used as input of the k-means clustering (with k=3). Eventually, pixels belonging to the most intense cluster were picked as the final binary mask.

2.4 Extraction of individual nuclei

The goal of the second step of the segmentation was assigning unique labels to each nuclei. In particular, it consisted of separating larger clusters into smaller ones if the clusters were connected artificially as a result of previous image processing. The task is often solved using the watershed algorithm [9]. However, whilst it is relatively easy to tune the algorithm to correctly divide a single cluster of nuclei, it is practically impossible to choose a set of parameters that works for every nuclei, region and image. Thus, again, we decided to use the model ensemble approach [10]. We generated a set of watershed segmentation results for each image with 5 different local maxima search region and 4 various minimal distances between markers. Then, for each resulting segmentation mask we performed binary edge extraction and stored results as an ensemble of binary edge masks. Next, we extracted the most probable nuclei edges by choosing edges present in at least 80% of edge masks. The resultant binary edge mask was superimposed over the most probable binary mask obtained in the previous step. On this image, we eventually applied once more the watershed algorithm in order to isolate and label individual nuclei.

3 Results

We tested two setups of the algorithm. In the first case, we used an ensemble generated only by the KM, GMM and SDWFCM models. On the training set, this approach reached average DICE1 of 79.2% (from 67% on image01 to 86% on image10), average DICE2 of 54.8% (from 40% on image01 to 65% on image03) and the overall average DICE score of 67.0%. The result was up to a few percent better than the DICE scores obtained using ensembles based on single methods (KM - 65.9%, GMM - 64.0%, SDWFCM - 63.3%). In the second setup, we included masks generated using the DM model, which led to slightly improved overall results (DICE1 80.2%, DICE2 55.8%, average DICE 68.0%). However, as the DM model was learned on the entire training set, these results are likely overestimated. Nevertheless, the score reached by the second setup was somewhat higher than the DICE coefficient obtained using only DM (67.3%).

References

[1] Wang, Chang, Wu, Tsai and Sun. A color-based approach for automated segmentation in tumor tissue classification. In: Proc. Conf. of the IEEE Engineering in Medicine and Biology Society. pp.6577–6580, 2007.

[2] Magee, Treanor, Crellin, Shires, Smith, Mohee and Quirke. Colour Normalisation in Digital Histopathology Images. In: Proc. Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop), pp.100-111, 2009.

[3] Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers, Norwell, MA, USA. 1981.

[4] Zeng, Huang, Kang and Sang. Image segmentation using spectral clustering of Gaussian mixture models. Neurocomputing 144:346–356, 2014.

[5] Guo, Liu, Wu, Hong and Zhang. A New Spatial Fuzzy C-Means for Spatial Clustering. WSEAS Transactions on Computers 14:369-381, 2015.

[6] Dahl and Larsen. Learning Dictionaries of Discriminative Image Patches. In: Proc. British Machine Vision Conference. p.77, 2011.

[7] Veta and van Diest, Kornegoor, Huisman, Viergever and Pluim. Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images. PLoS One 8:e70221, 2013.

[8] Onan. On the Performance of Ensemble Learning for Automated Diagnosis of Breast Cancer. In: Artificial Intelligence Perspectives and Applications (AISC 347), pp.119-129, 2015.

[9] Beucher and Meyer. The morphological approach to segmentation: the watershed transformation. In: Mathematical Morphology in Image Processing, pp.433–481, 1993.

[10] Angulo and Jeulin. Stochastic watershed segmentation. In: Proc. 8th Int Symposium on Mathematical Morphology, pp.265–276, 2007.