Title: Analysis and Classification of Microscopy Images with Cell Border Distance Statistics

Introduction: In the following, we describe an image analysis approach that is able to indirectly capture cell size statistics and cell border structure. Furthermore, we show how these features can be used together machine learning techniques to categorize microscopy images.

Cell sizes and the structure of cell borders are often mentioned as relevant features by medical experts for the task of categorizing microscopy data of histopathological images ex vivo. The microscopic detection of characteristic pathological changes of cell morphology and tissue alterations out of histological slides obtained from tissue biopsies are, for instance, the basis for a pathologist to detect and classify a neoplastic tumor. However, quantifying these features in a consistent manner is a challenging task, which we tackled with an image analysis algorithm that will be described in the following. Our proposed algorithm consists of the following main steps:

1. Segmentation of cell borders,
2. Calculation of distance maps and quantization,
3. Classification with kernel support vector machines using the resulting histograms.

Materials and Methods: *(Algorithm Step 1)* Cell borders can be recognized in the microscopy images simply by bright edge pixels. However, due to the non-uniform illumination present during acquisition, global thresholding techniques are unable to find a single threshold that allows for separating the cell borders from the background. Therefore, we use local normalization by first computing the maximum and minimum intensity in a 27 x 27 pixel neighborhood. The intensity of the current pixel is then transformed by stretching the (local) min/max range to the full gray value range [0, 255]. The resulting image is further smoothed by a Gaussian filter of size 7 x 7. The binarization can now be performed on this transformed image with a global threshold, since all intensities have been already normalized locally. The threshold can be tuned either by visual inspection or by empirical analysis on a given dataset. The resulting binary image is further processed by skeletonization with the Zhang-Suen algorithm [1]. These simple operations already yield a sufficient quality of the cell border segmentation. As an alternative, we also tested and analyzed the method of [2] and especially the oriented filtering technique described therein. Our results showed that this method is not applicable when image noise disturbs local orientation estimation significantly.

(Algorithm Step 2) Obtaining statistics for cell sizes usually requires region segmentation algorithms or connected component analysis applied to the cell border results. Whereas, region segmentation requires homogeneous areas, the latter method is not robust with respect to errors of the cell border segmentation. Due to this reason, we obtained cell size statistics indirectly by using the distance transform on the binary cell border image. Values in the resulting distance map h are defined by the smallest distance of the pixel to a cell border pixel. A distance map with larger values indicates larger cell sizes. An important property of the distance map is its robustness with respect to missing parts of the cell borders. A connected component analysis in this case would always result in large merged cells that lead to biased statistics. In contrast, distance values are only marginally affected. However, false-positive cell border pixel could lead to severe changes of the distance values. In our case, we avoid a large portion of false detections by the skeletonization operation performed in the beginning. Given a distance map, we create a histogram with k bins and relative frequencies. The number of bins has to be tuned with respect to the performance on a held-out test set, since a large number would allow for capturing fine-grained changes in the statistics, but a low number of bins might be more effective when only a few training examples are given for a latter training phase.

(Algorithm Step 3) Cell border statistics can be a very effective tool to categorize microscopy data into pre-defined classes. In our case, we used a support vector machine with Gaussian kernel as implemented in the `sklearn` framework. This classifier allows for non-linear separation between classes and can be also easily applied to multi-class scenarios with the one-vs-all principle. The kernel width hyperparameter as well as the trade-off parameter C can be estimated by 5-fold cross-validation.

(Materials and Experimental Setup) We applied our algorithm for the detection of primary head and neck cancer. Patient with suspicion of head and neck cancer underwent standard diagnostic endoscopy of the upper aerodigative tract. Before taking biopsies from suspicious areas and the neighboring normal-appearing mucosa to define the tumor extension, confocal laser endoscopy images were taken from these areas. Then, standard biopsies were taken exactly from the same spots and delivered for standard histopathology. As a consequence each confocal laser endoscopy image could be classified as showing tumorous or non-tumorous areas or both. The different areas were labeled in the images. Data and labels from images of normal and cancerous tissue of 11 patients

were used to automatically learn the differences in our cell border statistics. The analysis and performance evaluation was done with a leave-one-patient-out strategy to prevent the machine learning algorithms from learning a bias specific for single patients.

Results: We were able to achieve an accuracy of 74% and an area under the receiver operator characteristic curve of up to 92%, which shows that our compact statistical cell characteristics indeed contain valuable information for this challenging task. Figure 1 displays an example image used during our experiments along with the results of the cell border detection and the calculation of cell border distance statistics.

Summary: We briefly presented a simple image analysis algorithm that allows for classifying microscopy images using cell size statistics. A critical step of the method is still the cell border segmentation, which needs to be adapted for a given task. If images with manually labeled cell borders would be available, a model for the cell border appearance could be learned with the techniques described in [3].

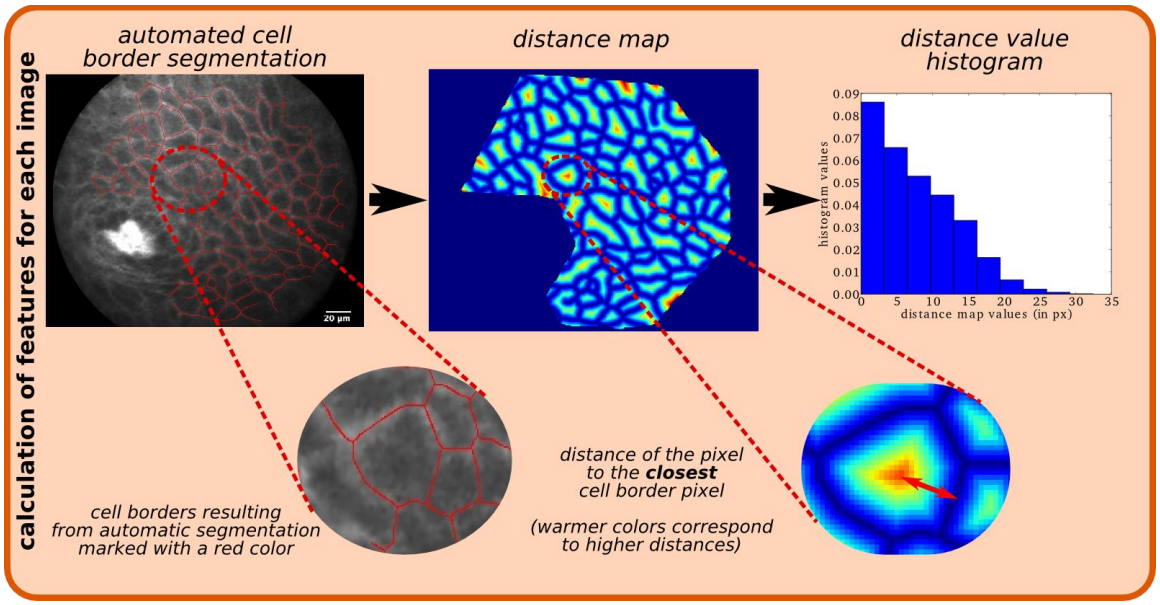

Fig. 1 Example results of the cell border segmentation and calculation of distance statistics

Literature

[1] TY Zhang and Ching Y. Suen. A fast parallel algorithm for thinning digital patterns. Communications of the ACM, 27(3):236–239, 1984.

[2] Herbert Süße, Wolfgang Ortmann, Christian Lautenschläger, Marco Körner, Carsten Schmidt, Andreas Stallmach, and Joachim Denzler. Oriented differences of boxes operators for blood vessel segmentation and analysis in confocal laser endomicroscopy images with minimal user interaction. In Proceedings of the 17th Annual Conference in Medical Image Understanding and Analysis (MIUA), pages 155–160, 2013.

[3] Clemens-Alexander Brust, Sven Sickert, Marcel Simon, Erik Rodner, and Joachim Denzler. Convolutional patch networks with spatial prior for road detection and urban scene understanding. In Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP), 2015. accepted for publication, pre-print available at http://hera.inf-cv.uni-jena.de:6680/pdf/Brust15:CPN.pdf.