AUTOMATIC IDENTIFICATION OF RELEVANT TISSUE SECTIONS IN HISTOLOGICAL WHOLE-SLIDE IMAGES

André Homeyer, Andrea Schenk, Heinz-Otto Peitgen
Fraunhofer MEVIS, Universitätsallee 29, 28359 Bremen, Germany
E-mail: andre.homeyer@mevis.fraunhofer.de

Uta Dahmen, Olaf Dirsch, Hao Jin
Jena University Hospital, Jena, Germany

KEY WORDS: Virtual Microscopy, Whole-Slide Imaging, Image Analysis, Histology

1. INTRODUCTION
With the rise of Virtual Microscopy, automatic image analysis has great potential to improve the speed and accuracy of pathological investigations [1]. For this purpose, it has to be able to cope with the inherent difficulties of histology. Firstly, histological slides are frequently affected by different kinds of artifacts, such as air entrapments and contaminations with dirt. Secondly, it is common practice to place multiple sections, often of different tissue types, on one slide. We present a fast and robust image analysis algorithm for the automatic separation of relevant tissue sections from irrelevant tissue sections and artifacts, and demonstrate its use in a practical software for the determination of tissue parameters.

2. METHODS
Initially, the entire slide is divided into small quadratic regions. A Random Forest classifier [2] then considers spectral and textural features at different magnifications in order to classify each region as either "relevant tissue", "irrelevant tissue" or "background". Eventually, locally-connected non-background regions which comply to certain constraints are merged to form the final sections. The algorithm can be easily adapted to different stainings by retraining the Random Forest classifier. For this purpose, our software implementation provides a special training mode where the user can pan and zoom through the whole-slide image in order to select representative examples of the three region classes.

3. RESULTS
The algorithm was evaluated on 51 HE- or Brdu-stained whole-slide images with sizes between 0.7 and 5.1 gigapixels. Each image contained 2–4 rat liver sections and up to 3 rat intestine sections that had to be sorted out (Fig. 1). Especially the Brdu-stained slides exhibited severe air and dirt artifacts. Of the 144 liver tissue sections, 143 were identified with a dice coefficient of more than 0.96. All intestine tissue sections and dirt artifacts were sorted out. In addition to being robust and accurate, the algorithm proved to be fast. On a standard notebook computer, the algorithm processed each image in 14–98 seconds.