

Robust extraction of quantitative structural and textural information from histological images

Q. Caudron^{1,*}, R. Garnier¹, K. A. Watt², J. G. Pilkington², J. M. Pemberton², T. A. Aboellail³, B. T. Grenfell^{1,4}, A. L. Graham¹,

1 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

2 Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3FL, UK

3 Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO, USA

4 Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

*** E-mail: qcaudron@princeton.edu**

Abstract

Introduction

Histopathology is widely practiced in medical and veterinary fields, where it is an essential tool in the diagnosis and discovery of infections, cancers, damaged tissue, and other conditions. In clinical settings, interpretation of histological samples is typically performed via microscopy by a pathologist. This results in a descriptive analysis of several aspects of the tissue's presentation, such as nuclear density, general cell morphology, or the presence of tumours. This time-intensive and expensive process is sensitive to intra- and inter-operator variation and subjectivity in interpretation. In addition, information in pathology reports is qualitative or semi-quantitative, which restricts the use of statistical tools for classification or quantitative analysis.

Recent advances in imaging technology and computational power have led to the emergence of a quantitative approach to histopathological analysis. Computer vision algorithms are applied to digitised histological preparations, extracting quantitative information about structural features in the tissue. These methods have been successfully applied to the high-throughput analysis of tissue banks for disease classification and phenotyping (Peng2008), and, increasingly, to aid in the detection and identification of disease in hospital settings (Tang2009). Computer-aided diagnosis combines image processing and artificial intelligence methods to aid pathologists in detecting and quantifying irregularities in the tissue, and is currently demonstrating extremely promising results in the fields of diagnostic and intervention radiology.

In both high-throughput and laboratory settings, a certain level of image quality is almost guaranteed, allowing image processing algorithms to successfully and reproducibly identify structures of interest in digital images. However, little progress has been made in applying these methods to field ecology, where on-site conditions often constrain the state of tissue samples or capture of images.

Here, we demonstrate that quantitative information can be extracted from high-variance images in a robust manner. We apply our algorithms to measures of liver state in the Soay sheep, from 4306 images of histological slides prepared from the liver necropsies of 141 individuals. We begin by presenting the Soay sheep system and associated data. We then describe the image capture process and preprocessing applied to images with the aim of reducing some of the variance in their colour, luminosity, and contrast. Then, we present various measures of structure and texture, which we compare with a traditional pathologist's analysis of the slides. We finish by comparing these measures to a traditional histological analysis of these slides by pathologist grading.

Methods

Soay Sheep

Background on the Soay.

The data we have : physical measurements, necropsies.

Histology preparation (slice, stain, slides per individual, ...)

Image Acquisition and Processing

Imaging was performed on a Nikon Eclipse 80i bright-field microscope with a Nikon CFI Plan Achromat 4x objective. Fields of view measuring 2×3 mm were captured in RAW format on a Canon 600D SLR camera, attached to the microscopy with an Amscope DSLR adaptor. For each image, the field of view and focus were found manually. Up to five 18-megapixel images were captured from each slide, at 14 bits per channel, in the three-channel sRGB colour space. Fields of view were selected to ensure they did not contain tears, folds, or other imperfections in the tissue caused by sample preparation.

Images were treated with a preprocessing stage upon capture. Saturation was increased globally, but also preferentially in low-saturation pixels. Yellows were reduced, to compensate for the microscope's lamp, and luminosity was increased in the highlights in order to improve separation of the tissue from the background. Finally, local midtone contrast was enhanced to increase sharpness without amplifying noise.

After acquisition and preprocessing, images were processed in Python with the scikit-image package. Images were first sigmoid-corrected for contrast, and converted to relative luminance greyscale. Thresholding was performed adaptively, where the threshold for a given pixel is defined as the Gaussian-weighted mean of a neighbourhood of 150 pixels. Finally, connected components under 100 pixels in size were removed.

Measures

- 1 Gabor filter directionality
- 2 Gabor filter scale
- NOTE GABOR - phase-insensitive
- 3 normalised Lacunarity
- 4 Shannon entropy
- 5 Deconvolution foci count and size
- 6 Tissue to sinusoid ratio

Results

Measure Distributions

Plots of distributions (include bighorn ?)

Intra- and inter-individual variances

PCA

Pathologist Analysis

Correlation with image measures

Discussion

Acknowledgements

QC and BTG were supported by funding from the US Department of Homeland Security contract HSHQDC-12-C-00058. BTG acknowledges support from the Bill & Melinda Gates Foundation. BTG was funded by the RAPIDD program of the Science and Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health.

References

Figure Legends

Tables

Locality	Population	Birth rate	τ
Bornholm	47100	19.4	15
Faroe Islands	28200	29.4	15
Reykjavík	47100	24.1	18
Hafnarfjörður	6000	22.4	8
Akureyri	7000	22.7	19
Vestmannaeyjar	3600	23.5	7

Table 1. Mean population sizes, birth rates, and sensitivity thresholds τ for each locality. Population sizes and annual birth rates per thousand are given as the mean over the study period. Thresholds were fit by maximising the correlation between the mean simulated epidemic time-series and the reported incidence data.

Figures

Figure 1. Reported and predicted biweekly incidence for Bornholm, the Faroe Islands, and four localities in Iceland. The observed data is in blue. For the predicted time-series, the mean value of incidence simulations is plotted as a dark red line, with 95% confidence intervals given in light red. Bornholm : $R^2 = 0.78$; Faroe Islands : $R^2 = 0.55$; Reykjavík : $R^2 = 0.73$; Hafnarfjörður : $R^2 = 0.86$; Akureyri : $R^2 = 0.80$; Vestmannaeyjar : $R^2 = 0.77$.

Figure 2. Reporting rates and seasonalities. Seasonality is plotted as a function of the biweek, with 95% confidence intervals in light blue.

Figure 3. Predictability of epidemic sizes. The mean predicted size of each epidemic as a function of its observed size, from ten thousand simulations. Red lines are the regression lines with the following coefficients of determination and slopes – Bornholm : $R^2 = 0.76$, gradient = 1.07; Faroe Islands : $R^2 = 0.77$, gradient = 0.60; Reykjavik : $R^2 = 0.64$, gradient = 0.96; Hafnarfjörður : $R^2 = 0.88$, gradient = 1.18; Akureyri : $R^2 = 0.49$, gradient = 0.72; Vestmannaeyjar : $R^2 = 0.76$, gradient = 1.23. The green line is the zero-intercept, gradient-one line representing a one-to-one match between observation and prediction.