A machine learning approach for the automated segmentation and tracking of bacteria in time-lapse microscopy

Johan H. van Heerden, Coco van Boxtel, Anne Doerr, Frank J. Bruggeman

Background

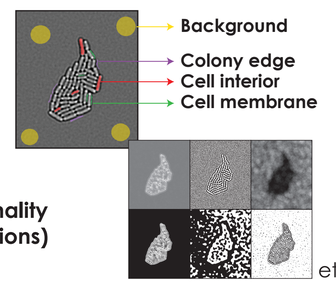
It is becoming increasingly apparent that the characteristics of individual cells can deviate substantially from population averages. Microscopic methods provide an essential means for studying heterogeneity in microbial population. However, the automated and accurate identification of growing bacterial cells in microscopic images is surprisingly challenging. Many methods are available, but these often require specific optical configurations or the use of fluorescent markers, and generally involve significant user intervention to achieve good segmentation results.

Here we outline a segmentation strategy that exploits the power and flexibility of machine learning to identify and track individual bacterial cells growing into dense microcolonies. The use of machine learning offers a generic solution that can easily be tailored for individual optical configurations and different bacterial morphologies.

1 Model training & classification

TRAINING DATA

Define class examples



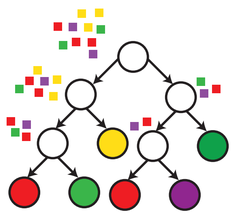
Examples of known pixel classes are defined.

Increase dimensionality (image transformations)

Several image transformations are applied (typically > 100) to generate many unique representations of each class example. Importantly, spatial information is included.

Train model (Random forest)

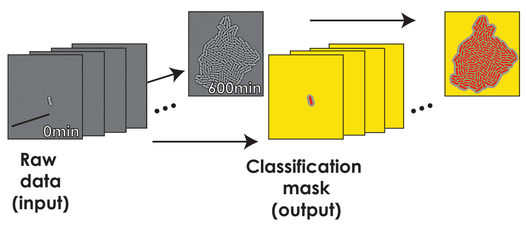
Using pixel class examples, find thresholds that optimally distinguishes classes



Generate a classification model from traning data, using a Random forest algorithm. Typically a single frame from a timelapse experiment is used.

EXPERIMENTAL DATA

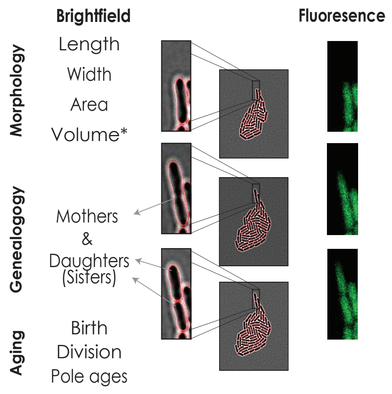
Apply Model



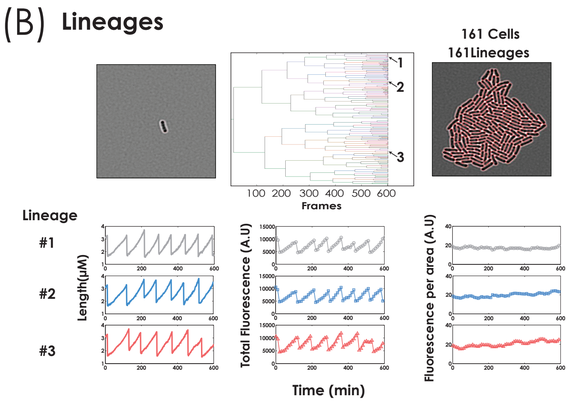
Classify unknown pixels using generated random forest.
Each pixel is tested against 300 trees and the consensus class is assigned.

3 Feature extraction & analysis

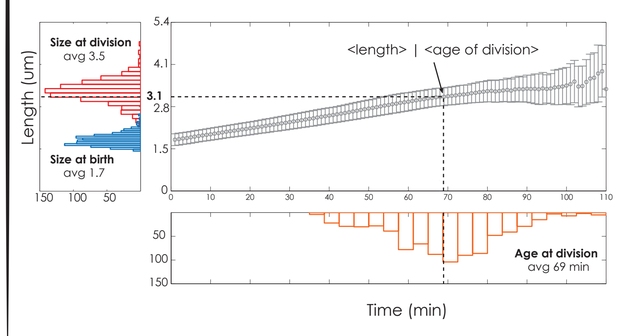
(A) **Single cell features**



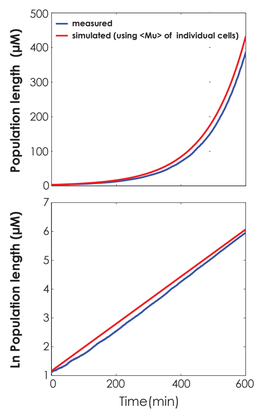
(B) **Lineages**



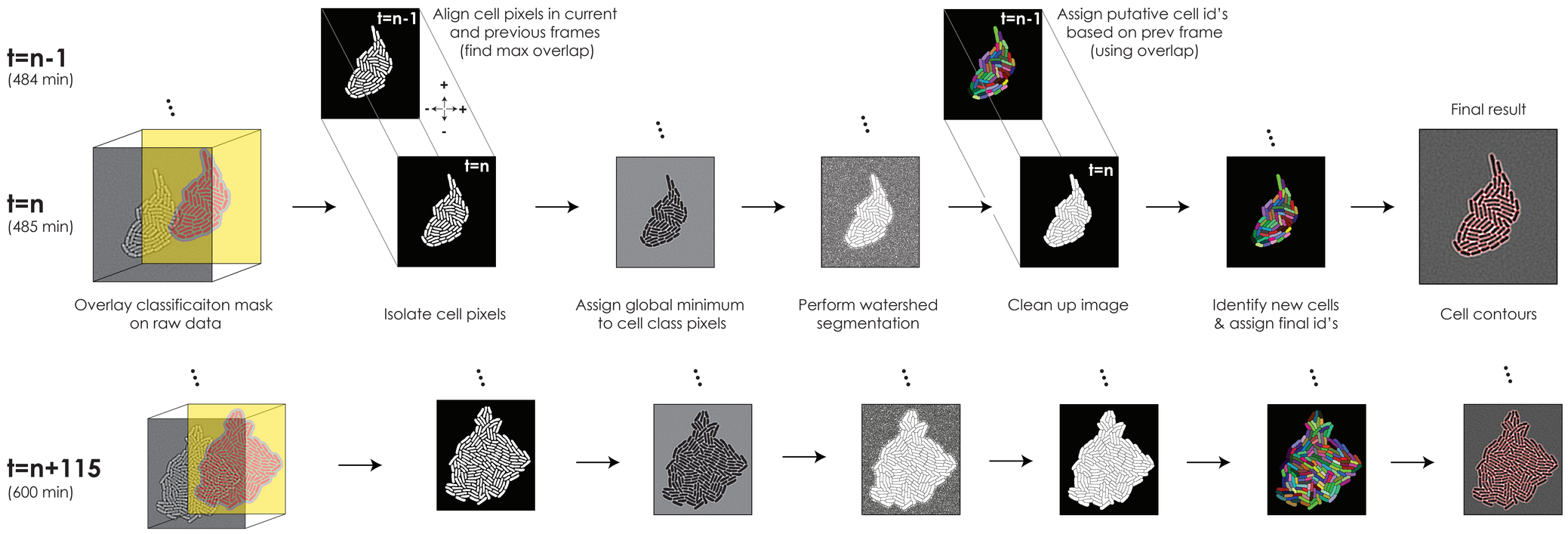
(C) **Statistics**



(D) **Population-level characteristics**



2 Segmentation & tracking



Concluding remarks

The initial classification of pixels into different classes proves to be an effective means for discriminating cell objects in images. This step is used as a very good first "guess", which is further refined using basic morphological operations and standard watershed segmentation in combination with segmentation information from the preceding frame in a time-lapse series. Our pipeline includes extensive feature extraction that allows for the detailed analysis of growth and gene expression dynamics. We have tested our method on brightfield time-lapse data of growing *E. coli* and *B. subtilis* cultures, and can track single cells growing into microcolonies for 7 to 10 generations, with a high degree of accuracy and no need for user intervention.

VU VRIJE UNIVERSITEIT AMSTERDAM
NWO
WEKA The University of Waikato
MATLAB