



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

**SCHOOL OF SCIENCES
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS**

**POSTGRADUATE PROGRAM
"INFORMATION TECHNOLOGIES IN MEDICINE AND BIOLOGY"**

MASTER THESIS

Bacterial image analysis based on time-lapse microscopy

Athanasios D. Balomenos

SUPERVISOR: Elias Manolakos, Associate Professor
COSUPERVISOR: Panagiotis Tsakanikas, Postdoctoral Researcher

ATHENS

AUGUST 2013



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Επεξεργασία βακτηριακής εικόνας από time-lapse
μικροσκοπία**

Αθανάσιος Δ. Μπαλωμένος

ΕΠΙΒΛΕΠΩΝ: Ηλίας Μανωλάκος, Αναπληρωτής Καθηγητής
ΣΥΝΕΠΙΒΛΕΠΩΝ: Παναγιώτης Τσακανίκας, Μεταδιδακτορικός Ερευνητής

ΑΘΗΝΑ

ΑΥΓΟΥΣΤΟΣ 2013

MASTER THESIS

Bacterial image analysis based on time-lapse microscopy

Athanasios D. Balomenos

R.N.: PIV084

SUPERVISOR: **Elias Manolakos**, Associate Professor
COSUPERVISOR: **Panagiotis Tsakanikas**, Postdoctoral Researcher

EXAMINING COMMITTEE: **Katia Karalis**, Investigator - Associate Professor Level,
Biomedical Research Foundation of the Academy of Athens
Elias Manolakos, Associate Professor
Alexandros Eleftheriadis, Associate Professor

August 2013

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Επεξεργασία βακτηριακής εικόνας από time-lapse μικροσκοπία

Αθανάσιος Δ. Μπαλωμένος

A.M.:ΠΙΒ084

ΕΠΙΒΛΕΠΩΝ: **Ηλίας Μανωλάκος**, Αναπληρωτής Καθηγητής
ΣΥΝΕΠΙΒΛΕΠΩΝ: **Παναγιώτης Τσακανίκας**, Μεταδιδακτορικός Ερευνητής

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: **Κάτια Καραλή**, Ερευνήτρια Β', Ίδρυμα Ιατροβιολογικών
Ερευνών της Ακαδημίας Αθηνών (ΙΙΒΕΑΑ)
Ηλίας Μανωλάκος, Αναπληρωτής Καθηγητής,
Πανεπιστήμιο Αθηνών
Αλέξανδρος Ελευθεριάδης, Αναπληρωτής Καθηγητής,
Πανεπιστήμιο Αθηνών

Αύγουστος 2013

ABSTRACT

Time-lapse microscopy now enables detailed imaging data generation and monitoring of dynamic cellular processes at the single cell level. Recent studies have highlighted the usage and importance of this technology for investigating biological noise in gene regulation, cell growth and proliferation etc. Mathematical and statistical model development is of growing interest in capturing and testing hypothesis regarding the dynamical behavior of biological systems. Modeling bacterial communities forming biofilms relies on the efficient and accurate extraction of information from time-lapse microscopy data (image frame sequences) of growing bacterial colonies. However, the analysis of such "cell movies" data is currently very time consuming and error prone since it is essentially performed by human-experts. In this thesis we address this important limitation in a multi-resolution image analysis framework.

We have developed a methodology for identifying accurately the boundaries of individual bacterial cells and tracking them from frame to frame so as to construct the cells' genealogy (bacterial cell segmentation and lineage tree construction) even in large-size microbial communities where there is great difficulty in identifying the individual cell boundaries. The automated and novel pipeline of algorithms we have developed combines methods from image processing and machine learning to segment and track bacteria precisely.

The pipeline has been tested and evaluated with two different cell movies datasets and several images produced by different labs. The developed methodology has been shown to achieve high F-measure score (above 95%) in each evaluation case. It can be applied to different image modalities, such as phase contrast, bright field, and fluorescent, produced by optical and confocal microscopy. Using extensive experimentation we demonstrate the robustness and reliability of the proposed pipeline regardless of the image modality used. Our image processing pipeline is fully automated, computationally efficient and suitable for high throughput analysis of bacterial cell movies without any human intervention on its calibration.

SUBJECT AREA: Image Analysis and Machine Learning

KEYWORDS: bacterial segmentation, cell counting, cell lineage construction, cell feature extraction and visualization, expectation-maximization.

ΠΕΡΙΛΗΨΗ

Η time-lapse μικροσκοπία επιτρέπει πλέον τη λεπτομερή δημιουργία δεδομένων από δυναμικές κυτταρικές διεργασίες σε επίπεδο μεμονωμένων κυττάρων (single cell level). Πρόσφατες μελέτες έχουν τονίσει τη χρήση και τη σημασία αυτής της τεχνολογίας για τη διερεύνηση του βιολογικού θορύβου στη ρύθμιση των γονιδίων, την ανάπτυξη και τον πολλαπλασιασμό κυττάρων κλπ. Τα μαθηματικά και τα στατιστικά μοντέλα παρουσιάζουν αυξανόμενο ενδιαφέρον για τη σύλληψη και τον έλεγχο υποθέσεων σχετικά με τη δυναμική συμπεριφορά των βιολογικών συστημάτων. Για τις βακτηριακές κοινότητες που σχηματίζουν βιοϋμένια αυτό εξαρτάται από την αποτελεσματική και ακριβή άντληση πληροφοριών από δεδομένα time-lapse μικροσκοπίας (ακολουθίες εικόνων) αναπτυσσόμενων βακτηριακών αποικιών. Ωστόσο, η ανάλυση τέτοιων δεδομένων «κυτταρικών ταινιών» σήμερα είναι πολύ χρονοβόρα και επιρρεπής σε λάθη, δεδομένου ότι ουσιαστικά πραγματοποιείται από έναν άνθρωπο-εμπειρογνώμονα. Στην παρούσα εργασία ασχολούμαστε με αυτό το σημαντικό περιορισμό σε ένα πολλαπλής ανάλυσης πλαίσιο επεξεργασίας εικόνας.

Έχουμε αναπτύξει μια μεθοδολογία που προσδιορίζει με ακρίβεια τα όρια μεμονωμένων βακτηριακών κυττάρων, τα ανιχνεύει από εικόνα σε εικόνα και κατασκευάζει τη γενεαλογία τους (κατάτμηση βακτηριακών κυττάρων και κατασκευή γενεαλογικών δέντρων) ακόμη και σε μεγάλες μικροβιακές κοινότητες όπου υπάρχει μεγάλη δυσκολία στον προσδιορισμό των επιμέρους ορίων των κυττάρων. Η αυτοματοποιημένη και καινοτόμος σωλήνωση (pipeline) των αλγορίθμων που αναπτύξαμε συνδυάζει μεθόδους από την επεξεργασία εικόνας και τη μηχανική μάθηση για την ακριβή βακτηριακή κατάτμηση και ανίχνευση.

Η προτεινόμενη μεθοδολογία έχει δοκιμαστεί και αξιολογηθεί με δύο διαφορετικά σύνολα δεδομένων κυτταρικών ταινιών και αρκετές ανεξάρτητες εικόνες από διαφορετικά εργαστήρια. Έχουμε αποδείξει ότι η μέθοδος που αναπτύχθηκε επιτυγχάνει υψηλή βαθμολογία F-measure (άνω του 95%) σε κάθε περίπτωση αξιολόγησης. Επίσης το λογισμικό μπορεί να εφαρμοστεί σε διαφορετικούς τύπους εικόνας, όπως phase-contrast, bright field, και φθορίζουσας (fluorescent), που προέρχονται τόσο από οπτική όσο και από συνεστιακή μικροσκοπία. Πραγματοποιώντας εκτεταμένα πειράματα που αποδεικνύουν την ευρωστία και την αξιοπιστία της προτεινόμενης σωλήνωσης, ανεξάρτητα από το τύπο της εικόνας που χρησιμοποιείται. Η μεθοδολογία επεξεργασίας εικόνας, που αναπτύξαμε, είναι πλήρως αυτοματοποιημένη, υπολογιστικά αποδοτική και κατάλληλη για υψηλής ρυθμαπόδοσης ανάλυση βακτηριακών κυτταρικών ταινιών χωρίς καμία ανθρώπινη παρέμβαση κατά τη βαθμονόμηση της.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Ανάλυση Εικόνας και Μηχανική Μάθηση

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Βακτηριακή κατάτμηση, απαρίθμηση κυττάρων, δημιουργία κυτταρικής γενεαλογίας, εξαγωγή και οπτικοποίηση βακτηριακών χαρακτηριστικών, expectation-maximization.

ACKNOWLEDGMENTS

First of all I would like to thank my advisor Dr. Elias S. Manolakos, Assoc. Professor, University of Athens. His influence, guidance and insights helped me explore the extremely interesting field of biomedical image processing and complete successfully this challenging thesis. I would also like to thank him for trusting on me this very interesting and promising project, and for suggesting the collaboration with his ex-student and one of the experts in the image analysis field, Dr. Panos Tsakanikas. His contribution however, is much more than that, as he has contributed significantly in the formation of new research goals for my future scientific career.

I would like to thank also Dr. Panagiotis Tsakanikas of the Biomedical Research Foundation of the Academy of Athens, for his guidance through many intricate image processing mechanisms and for providing all the support and expertise needed for completing this study. I would also like to thank him for the ideal conditions he has created throughout our collaboration.

I would like to thank Dr. Konstantinos Koutsoumanis, Assistant Professor, Department of Food Science and Technology, School of Agriculture, Aristotle University of Thessaloniki, for his guidance and valuable comments, which proved to be essential in order for this thesis to be relevant to the applied microbiology community.

I am also thankful to Dr. Katia Karalis, of the Biomedical Research Foundation of the Academy of Athens and Dr. Alexandros Eleftheriadis, Assoc. Professor, University of Athens, for their participation in the examining committee and useful comments that helped me improve the presentation of this work.

Finally, I want to thank my family and my friends for their continuous support, patience and love.

Athanasios D. Balomenos

Athens, August 2013

TABLE OF CONTENTS

FOREWORD.....	14
1. INTRODUCTION	15
1.1 From Single Cells to Micro-colonies and Biofilms	15
1.2 Microscopy Types	16
1.2.1 Optical Microscopy	16
1.2.2 Confocal Laser Scanning Microscopy	21
1.3 Necessity for Software Development.....	23
1.4 Thesis Outline	24
2. STATE OF THE ART.....	25
2.1 Existing Software Packages	25
2.2 Developed Methodology	26
3. METHODS.....	28
3.1 Preprocessing	28
3.2 Colonies Segmentation	29
3.3 Single cell Segmentation Algorithm	29
3.3.1 Collinear Object Processing	31
3.3.2 Gaussian Mixture Modeling.....	33
3.3.3 Complex Object Processing.....	33
3.4 Cell Lineage Construction.....	38
3.5 Feature Extraction & Visualization	40
4. DATASETS & EVALUATION SCHEME.....	41
4.1 Datasets' Description	41
4.2 Evaluation Scheme.....	45
5. RESULTS & DISCUSSION.....	47

5.1 Pipeline Evaluation based on Ground truth 47

5.2 Software Evaluation versus State-of-the-Art..... 50

5.3 Discussion and Additional Software Capabilities..... 52

6. CONCLUSION AND FUTURE WORK..... 59

REFERENCES..... 60

TABLE OF FIGURES

Figure 1: Microscopic bright field image (last frame) of a growing micro-colony of <i>Bacillus megaterium</i> (left) and Microscopic bright field image with overlaid fluorescence channel of a growing micro-colony of <i>Bacillus megaterium</i> [12].	18
Figure 2: Microscopic phase contrast image of a growing micro-colony of <i>Escherichia coli</i> [13].	19
Figure 3: By attaching fluorescent proteins to the genetic circuit responsible for <i>B. subtilis</i> 's stress response we can observe the cells' pulses as green flashes [14].	20
Figure 4: Gene expression is inherently noisy [15]. These cells express two fluorescent proteins, one shown in red, and the other in green. Both genes are controlled by the same promoters; therefore, if gene expression were deterministic, all cells would have equal amounts of red and green (and therefore appear yellow). The presence of cells that are more red than green, or vice versa, results from stochastic fluctuations (noise) in gene expression.	21
Figure 5: <i>S. Typhimurium</i> micro-colony monitored by DIC microscope.	22
Figure 6: Micro-colony was observed [16]: (A) on the xy plane with phase-contrast microscopy; (B) on the xy plane with CLSM; (C) on the xz plane along the diameter-line "D" with CLSM.	23
Figure 7: Proposed pipeline overview.	28
Figure 8: Input image in preprocessing stage (left). Output image (right) after denoising and CLAHE application.	29
Figure 9: Input image in colonies segmentation stage (left). Output image (right), a mask with segmented colonies.	30
Figure 10: Initial segmentation image (bottom) created by multiplication of the two masks (top).	30
Figure 11: Skeletonization of three objects. The object of the top panel (green) has no junctions, i.e. collinear object, while the objects in the rest of panels (red) have more than one junction, i.e. complex object.	31
Figure 12: Bacterial fission [33]. In stage 5 the cell enters into the division stage. Our method detects objects having the shape of stage 5, i.e. bow tie shape.	32
Figure 13: Pipeline of identification of bowties in collinear objects	32
Figure 14: Schematic overview of the proposed segmentation methodology. (1) Colony segmentation, (2) object extraction, (3) skeleton based classification; (4) Complex object segmentation: 4a) Segmentation by watershed algorithm, 4b) "puzzle solving" step, 4c) dataset generation, 4d) unsupervised mixture modeling, 4e) final result of bacterial segmentation; (5) Collinear object segmentation: 5a) Application of "deep" valley criterion and identification of "bow ties", 5b) final result of bacterial segmentation. At step (3), there is a bifurcation leading to two different processing routes, according to the object type that is going to be segmented.	34
Figure 15: Dataset generation examples. The dataset's size is proportional to the object size.	35
Figure 16 Panel A: Representation of an image object in 3D space, the pixels intensities lie in the z axis. Panel B: Representation of the same object distance transform in 3D space, the distance values lie in the z axis. It is obvious that distance transform	

smoothes object abnormalities while sharpens the valley between the two bacteria comprising the object.....	36
Figure 17: Application of EM algorithm with MML criterion for model selection. Top: No center reduction. Bottom: The centers are reduced from 17 to15.	37
Figure 18: Lineage construction algorithm overview. (1) Segmented bacteria of two consecutive frames, (2) Bacterial contours and centroids extraction from frames n-1 and n respectively, and(3) Bacteria matching.....	39
Figure 19: Possible lineage sub-tree patterns. (1) Error free pattern, (2) False positive pattern (red rectangle) and correction (green rectangle) by merging, and (3) False negative pattern (red rectangle) and correction (green rectangles) by splitting.	40
Figure 20: Images' stack of SalPhase dataset.	42
Figure 21: Images' stack of GFP confocal dataset.	42
Figure 22: Single frame provided by MT. E. coli bacteria forming small micro-colonies	43
Figure 23: Single frame provided by CellTracer. E.coli bacteria forming a single micro-colony[13].	43
Figure 24: Single frame provided by TLM-Tracker [12]. B. megaterium bacteria forming a single micro-colony.	44
Figure 25: Single frame from multi-salPhase dataset. S. typhimurium bacteria forming multiple micro-colonies	44
Figure 26: Example of a ground truth	46
Figure 27: Demonstration of evaluation and segmentation results of a single frame of the salPhase dataset by the proposed methodology.	46
Figure 28: Evaluation results for SalPhase dataset. Panel A: Evaluation results for the complete dataset and cumulative segmentation errors (FNs+ FPs) (please refer to Chapter4). Panel B: The same as in Panel A but for the confocal GFP dataset. As we can see, the developed method, while fully automated and user friendly, achieves a high F-measure score (above 97%). This measure is consistent regardless the image acquisition modality (i.e. optical, confocal, phase contrast /bright field/GFP staining. ...	48
Figure 29: Evaluation results presentation for SalPhase dataset. Row 1 from left to right: Original dataset (stack of phase contrast images), cell segmentation result on one colony (red ellipses indicate the contours of segmented cells), and bacterial area visualization of the same colony and histogram of the segmented bacterial areas. Row 2 from left to right: Original dataset (stack of confocal GFP images), cell segmentation result (red ellipses indicate the contours of segmented cells), GFP visualization of the colony on the initial image and histogram of the bacterial GFP quantity. We can observe from the figure that the developed methodology is efficient and robust while fully automated for a stack of images, usually exhibiting limited intra-variability in terms of focus and illumination.	48
Figure 30: Bacterial Growth curves for each colony in the salPhase dataset (line 1) and in the GFP confocal dataset (line 2), corresponding colony Area Growth curve, and bacterial length distribution. We can observe from the figure that the developed methodology embeds the capability of measuring several colony and single cell properties of high importance for further analysis.....	49
Figure 31: Comparison of the fittings for salPhase dataset counted manually and by the developed software for a micro-colony; Tables present the kinetic parameters of microbial growth estimated.....	50

Figure 32: Software comparison evaluation with the state-of-the-art. First two columns present the segmentation/cell detection results of the developed and MicrobeTracker. For the CellTracer and TLM Tracker we show no segmentation images since they could not operate with all different datasets. The third column summarizes the performance of each method for every different dataset. As we can see from the segmentation results (first column) and the relative F-measure ($\geq 96\%$ for all cases), the proposed method is robust for different imaging modalities (optical and confocal phase contrast, optical bright field) and for data acquired by different labs. Furthermore, if one inspects the segmented images, it is clear that the cells' contours are more reliable, leading to a more efficient and robust cell property estimation and GFP measurements.51

Figure 33: Relative cell position to the centroid of the colony (colony mapping).53

Figure 34: Area visualization, we can see the evolution of cell area through time and cell lineage simultaneously, one easily see the "critical" cell size just before cell division. ..54

Figure 35: GFP quantification visualization through time (GFP confocal dataset). The root node of the lineage is a pseudo node, i.e. they do not represent an actual bacterium, because the first frame of the dataset contains a micro-colony of four bacteria and not a single cell.54

Figure 36: Cell division time visualization (red branch shows that the division time of a cell is more than 60 minutes). The root node of the lineage is a pseudo node, i.e. they do not represent an actual bacterium, because the first frame of the dataset contains a micro-colony of four bacteria and not a single cell.55

Figure 37: Average bacterium area per generation according to lineage of Fig. 33 (left). Average bacterium division time per generation according to lineage of Fig. 35 (right).55

Figure 38: Multi-salPhase dataset analysis (frame 65 and 78). The four colonies merged with each other. The rightmost red frames illustrate the input images. The cyan and the purple boxes illustrate the evaluation statistics of each colony and each frame correspondingly.56

Figure 39: Computational time curves. Execution time of each frame (left). Segmentation time per bacterium of each frame (middle). Segmentation time of each frame (right).58

TABLE OF TABLES

Table 1: Software Packages Overview	26
---	----

FOREWORD

This study constitutes a very interesting, important and compulsory part of my postgraduate studies in the Master's Program "Information Technologies in Medicine and Biology", which is organized and administrated by the Department of Informatics and Telecommunications of the National and Kapodistrian University of Athens (UoA), in cooperation with the Technological Educational Institute (TEI) of Athens, and in collaboration with the Foundation for Biomedical Research of the Academy of Athens (BRFAA) and the Institute of Informatics and Telecommunications of National Centre for Scientific Research "Democritos". The Thesis was written in Athens under the supervision of Elias S. Manolakos, Associate Professor at the Department of Informatics, University of Athens, and Dr. Panagiotis Tsakanikas, postdoctoral researcher in the Biomedical Research Foundation of the Academy of Athens, in the final semester of my studies.

Our collaboration with Dr. Koutsoumanis, Assistant Professor, Department of Food Science and Technology, School of Agriculture, Aristotle University of Thessaloniki, and his lab members was of major importance in order for us to understand the details intricate of the image types we would have to deal with. Moreover, their contribution was crucial as they conducted many experiments and provided us with all the experimental data we needed to thoroughly test and evaluate the developed algorithms. This collaboration set the goals of this study, which are in close correlation with the research interests of Dr. Koutsoumanis and his group in data extraction for bacterial micro-colony formation.

1. INTRODUCTION

In this chapter we define the problem of bacterial image processing and the importance of it for systems biology approaches. On the first section, we introduce the reader to the concepts of micro-colony formation. In the second section, we summarize the fundamental principles of microbiological microscopy. Furthermore, we discuss the necessity for software development in such fields.

1.1 From Single Cells to Micro-colonies and Biofilms

For many years predictive microbiology deals with the development of deterministic models based on studies with large microbial populations. Traditional mathematical models describe the growth of microbial populations as a whole, without considering the individual cells. In “real life” however, contamination of foods with pathogens usually occurs at very low numbers and the probability that their multiplication results in an infectious level at the time of consumption depends greatly on the kinetics of the contaminating cells as well as on whether they are able to grow. Recently, the importance of stochastic models which are able to predict the effects of more “realistic” contamination events (low microbial numbers) in food safety has been stressed out. Thus, predictive microbiology studies have focused on monitoring microbial kinetics at a single cell level.

Individual cells within clonal microbial populations exhibit a remarkable phenotypic variation which refers to epigenetic sources of population variation that do not involve changes in the genome [1]. For example, the production of a specific protein in genetically identical cells in an essentially identical environment can differ among cells owing to stochastic fluctuations (or “noise”) during transcription and translation, leading to differences in protein levels. The observable consequence of the above variation is the behavioral noise of single cells including the noise in the division time.

The ability of microbial communities to adhere on surfaces or interfaces (e.g. air-liquid interface) and to form biofilms has been proved to be an effective driving force for microbial survival and persistence. In general, microbial life in the form of biofilms represents a special strategy that allows cells to: (a) tolerate conditions and treatments that would be lethal to non-attached, planktonic cells and to exhibit dedicated stress responses [2-3], (b) grow and exhibit enhanced survival through metabolic cooperativity [4-5] and (c) acquire new genetic traits through horizontal gene transfer [6]. The strategies through which sessile bacteria develop the aforementioned characteristics (e.g. existence of persistent cells, cellular adaptation to antibiotics or/and preservatives) are of special scientific importance, mainly due to the severity of problems associated with biofilms and in comparison with planktonic counterparts (i.e. free floating, constrained or immobilized [7]). Microbial life in the form of biofilms is encountered in many (if not all) ecosystems, including eukaryotic tissues, and both natural and manmade surfaces [8]. As a consequence, biofilm formation in the food chain can be considered as a major risk since adherent cells may at any time be detached and contaminate the final products. Given that the physiology of detached cells is highly affected by their previous life in biofilm micro colonies, investigating the effect of biofilm forming conditions and their impact on subsequent physiology of biofilm cells is of utmost importance for assessing the risk of growth of detached cells.

The formation of a biofilm is a complex process and a number of genes are implicated while potential underlying mechanisms for the attachment of bacteria on biotic e.g. fresh produce or abiotic surfaces, e.g., industrial equipment, include: (i) the presence of extracellular polymeric substances or fimbriae; (ii) bacterial cell surface hydrophobicity (charge); (iii) divalent cationic bridges; (iv) bacterial strains; (v) structure of the surface

(rough or smooth, intact or cut / injured); (iv) presence of nutrients at the edges of cuts or at injury sites of fresh produce. In the case of plant tissue, bacteria may be entrapped due to internalization (vascular or other type, such as irrigation water), or translocation and this may enhance their resistance to sanitizing agents. Studies on the formation, organization and development of biofilm populations, together with those on cell-to-cell signaling mechanisms in such communities, are nowadays based on the knowledge that biofilm cells show different growth rates and gene expression profiles compared to planktonic cells of the same microorganism [9-11].

1.2 Microscopy Types

1.2.1 Optical Microscopy

Optical or light microscopy involves passing visible light transmitted through or reflected from the sample through a single or multiple lenses to allow a magnified view of the sample. The resulting image can be detected directly by the eye, imaged on a photographic plate or captured digitally. The single lens with its attachments, or the system of lenses and imaging equipment, along with the appropriate lighting equipment, sample stage and support, form the basic light microscope. The most recent development is the digital microscope, which uses a CCD camera to focus on the exhibit of interest. The image is shown on a computer screen, so eye-pieces are somewhat unnecessary.

Limitations of standard optical microscopy (bright field microscopy) lie in three areas;

- The technique can only image dark or strongly refracting objects effectively.
- Diffraction limits resolution to approximately 0.2 micrometres.
- Out of focus light from points outside the focal plane reduces image clarity (blurry image).

Live cells, in general, lack sufficient contrast to be studied successfully, since the internal structures of the cell are colorless and transparent. The most common way to increase contrast is to stain the different structures with selective dyes, but this often involves killing and fixing the sample. Staining may also introduce artifacts, apparent structural details that are caused by the processing of the specimen, and are thus not a legitimate feature of the specimen.

These limitations have all been overcome to some extent by specific microscopy techniques that can non-invasively increase the contrast of the image. In general, these techniques make use of differences in the refractive index of cell structures. It is comparable to looking through a glass window: one (bright field microscopy) does not see the glass but merely the dirt on the glass. There is a difference, as glass is a denser material, and this creates a difference in phase of the light passing through. The human eye is not sensitive to this difference in phase, but clever optical solutions have been developed to change this difference in phase into a difference in light intensity.

In order to improve specimen contrast and/or highlight certain structures in samples, special techniques must be used. A huge selection of microscopy techniques is available to increase contrast or label a sample. In bacterial microscopy the most significant trans-illumination techniques used are bright field, phase contrast, fluorescence and differential interference contrast (DIC) microscopy.

1.2.1.1 Bright field microscopy

Bright field microscopy is the simplest of all the optical microscopy illumination techniques. Sample illumination is transmitted (i.e., illuminated from below and observed from above, i.e. upright microscope) white light and contrast in the sample is caused by absorbance of the transmitted light in dense areas of the sample. Bright field microscopy is the simplest of a range of techniques used for illumination of samples in light microscopes and its simplicity makes it a popular technique. The typical appearance of a bright field microscopy image is a dark sample on a bright background, hence the name.

Bright field microscopy typically has low contrast with most biological samples since the difference in light absorption is limited. Staining is often required to increase contrast, which nevertheless prevents the use of live cells in many situations. Bright field illumination is useful for samples which have an intrinsic color, for example chloroplasts in plant cells. Bright field microscopy is a standard light microscopy technique and therefore magnification is limited by the resolving power possible with the wavelength of visible light. Limitations include low contrast of most biological samples and low apparent resolution due to the blur of out of focus material. The simplicity of the technique and the minimal sample preparation required are significant advantages.

1.2.1.2 Phase Contrast Microscopy

Phase contrast microscopy is an optical microscopy technique that converts phase shifts in light passing through a transparent specimen into brightness changes in the image. Light phase shifts are invisible, but become visible when shown as brightness variations.

Phase contrast microscopy is particularly important in biology, as it reveals many cellular structures that are not visible with a bright field microscope. These structures were made visible to earlier microscopy users only by sample staining, which killed the cells. The phase contrast microscopy made it possible for biologists to study living cells and how they proliferate through cell division.

The nucleus in a cell for example will show up darker than the surrounding cytoplasm. Contrast is excellent; however it is not for use with thick objects. Frequently, a halo is formed even around small objects, which obscures details. The system consists of a circular annulus in the condenser, which produces a cone of light. This cone is superimposed on a similar sized ring within the phase-objective. Every objective has a different size ring, so for every objective another condenser setting has to be chosen. The ring in the objective has special optical properties: it, first of all, reduces the direct light in intensity, but more importantly, it creates an artificial phase difference of about a quarter wavelength. As the physical properties of this direct light have changed, interference with the diffracted light occurs, resulting in the phase contrast image. As mentioned earlier, one disadvantage of phase-contrast microscopy is halo formation (halo-light ring).

When light waves travels through a medium other than vacuum, interaction with the medium causes the wave amplitude and phase to change in a manner dependent on properties of the medium. Changes in amplitude (brightness) arise from the scattering and absorption of light, which is often wavelength dependent and may give rise to colors. Photographic equipment and the human eye are only sensitive to amplitude variations. Without special arrangements, phase changes are therefore invisible. Yet, often these changes in phase carry important information.

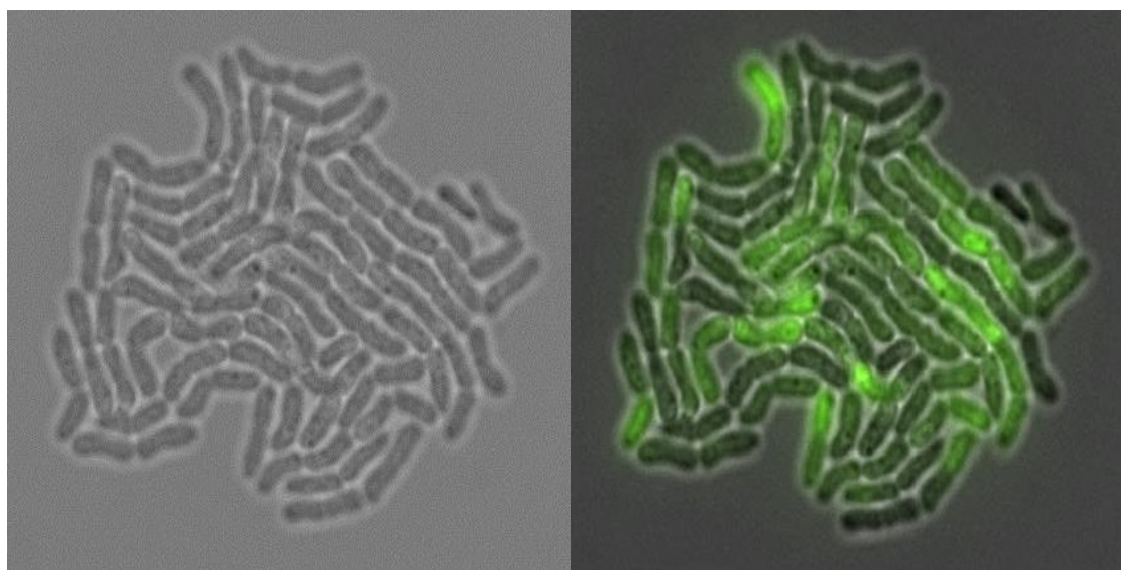


Figure 1: Microscopic bright field image (last frame) of a growing micro-colony of *Bacillus megaterium* (left) and Microscopic bright field image with overlaid fluorescence channel of a growing micro-colony of *Bacillus megaterium* [12].

Phase contrast microscopy proved to be such advancement in microscopy that its inventor Frits Zernike was awarded the Nobel Prize (physics) in 1953.

1.2.1.3 Fluorescence Microscopy

When certain compounds are illuminated with high energy light, they emit light of a different, lower frequency. This effect is known as fluorescence. Often specimens show their characteristic auto-fluorescence image, based on their chemical makeup.

This method is of critical importance in the modern life sciences, as it can be extremely sensitive, allowing the detection of single molecules. Many different fluorescent dyes can be used to stain different structures or chemical compounds. One particularly powerful method is the combination of antibodies coupled to a fluorophore as in immunostaining. Examples of commonly used fluorophores are fluorescein or rhodamine.

The antibodies can be tailor-made for a chemical compound. For example, one strategy often in use is the artificial production of proteins, based on the genetic code (DNA). These proteins can then be used to immunize rabbits, forming antibodies which bind to the protein. The antibodies are then coupled chemically to a fluorophore and used to trace the proteins in the cells under study.

Highly efficient fluorescent proteins such as the green fluorescent protein (GFP) have been developed using the molecular biology technique of gene fusion, a process that links the expression of the fluorescent compound to that of the target protein. This combined fluorescent protein is, in general, non-toxic to the organism and rarely interferes with the function of the protein under study. Genetically modified cells or organisms directly express the fluorescently tagged proteins, which enables the study of the function of the original protein *in vivo*.

Growth of protein crystals results in both protein and salt crystals. Both are colorless and microscopic. Recovery of the protein crystals requires imaging which can be done by the intrinsic fluorescence of the protein or by using transmission microscopy. Both

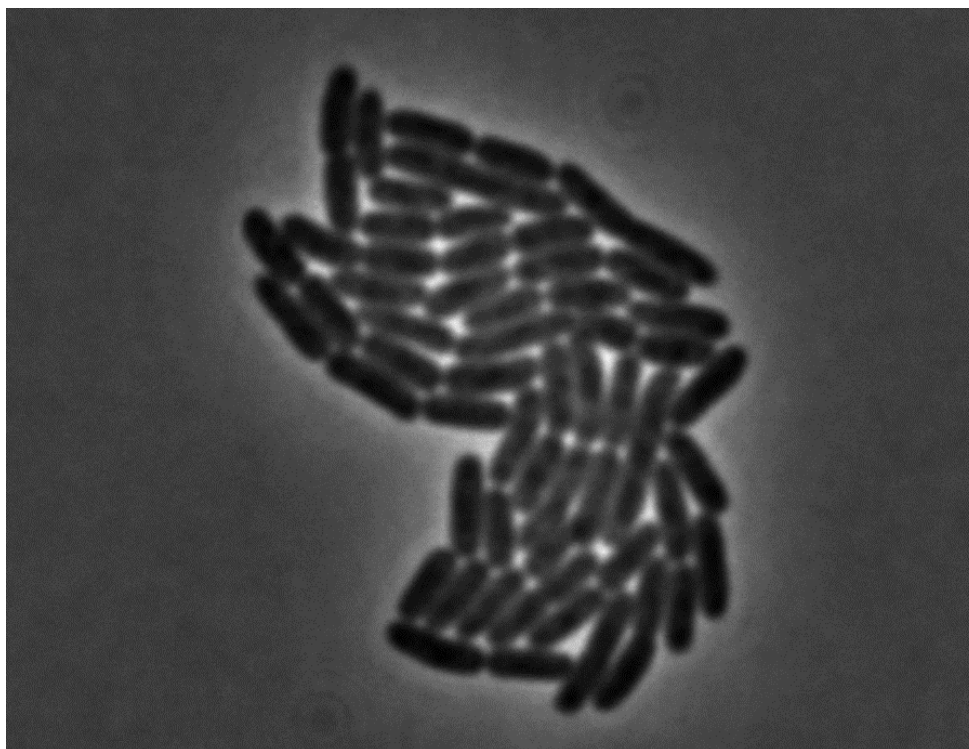


Figure 2: Microscopic phase contrast image of a growing micro-colony of Escherichia coli [13].

methods require an ultraviolet microscope as protein absorbs light at 280nm. Proteins will also fluoresce at approximately 353 nm when excited with 280 nm light.

Since fluorescence emission differs in wavelength (color) from the excitation light, an ideal fluorescent image shows only the structure of interest that was labeled with the fluorescent dye. This high specificity led to the widespread use of fluorescence light microscopy in biomedical research. Different fluorescent dyes can be used to stain different biological structures, which can then be detected simultaneously, while still being specific due to the individual color of the dye.

To block the excitation light from reaching the observer or the detector, filter sets of high quality are needed. These typically consist of an excitation filter selecting the range of excitation wavelengths, a dichroic mirror, and an emission filter blocking the excitation light. Most fluorescence microscopes are operated in the Epi-illumination mode (illumination and detection from one side of the sample) to further decrease the amount of excitation light entering the detector.

1.2.1.4 Differential Interference Contrast Microscopy

Differential interference contrast (DIC) microscopy, also known as Nomarski Interference Contrast (NIC) or Nomarski microscopy, is an optical microscopy illumination technique used to enhance the contrast in unstained, transparent samples. DIC works on the principle of interferometry to gain information about the optical path length of the sample, to see otherwise invisible features. A relatively complex lighting scheme produces an image with the object appearing black to white on a grey background. This image is similar to that obtained by phase contrast microscopy but without the bright diffraction halo.

DIC works by separating a polarized light source into two orthogonally polarized mutually coherent parts which are spatially displaced (sheared) at the sample plane, and recombined before observation. The interference of the two parts at recombination

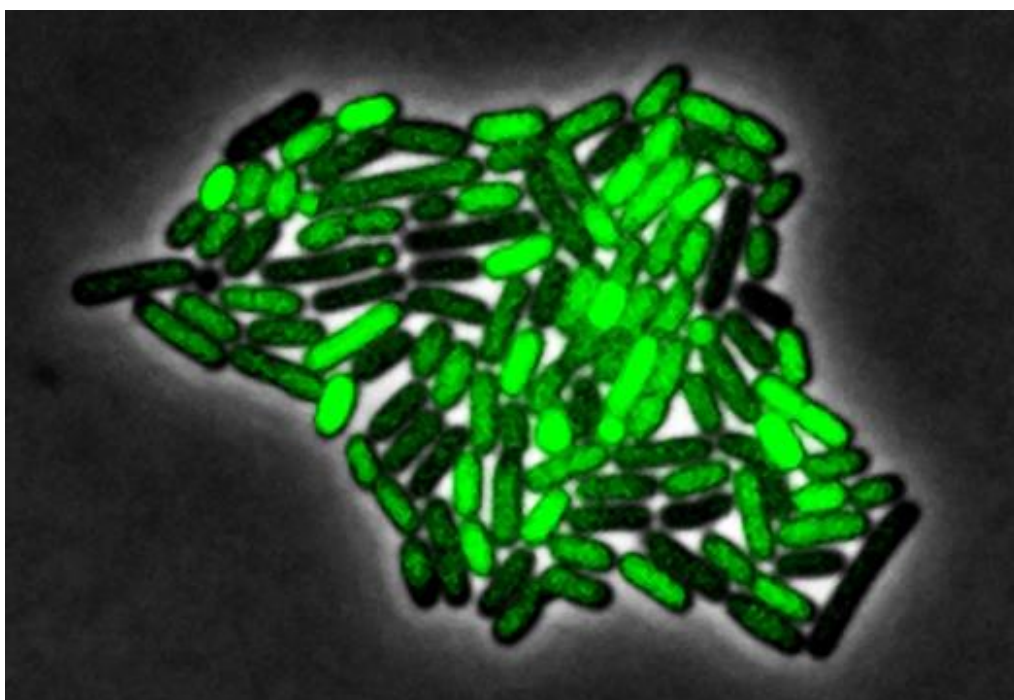


Figure 3: By attaching fluorescent proteins to the genetic circuit responsible for *B. subtilis*'s stress response we can observe the cells' pulses as green flashes [14].

is sensitive to their optical path difference (i.e. the product of refractive index and geometric path length). Adding an adjustable offset phase determining the interference at zero optical path difference in the sample, the contrast is proportional to the path length gradient along the shear direction, giving the appearance of a three-dimensional physical relief corresponding to the variation of optical density of the sample, emphasizing lines and edges though not providing a topographically accurate image.

DIC has strong advantages in uses, involving live and unstained biological samples, such as a smear from a tissue culture or individual water borne single-celled organisms. Its resolution and clarity in conditions such as this are unrivaled among standard optical microscopy techniques.

The main limitation of DIC is its requirement for a transparent sample of fairly similar refractive index to its surroundings. DIC is unsuitable (in biology) for thick samples, such as tissue slices, and highly pigmented cells. DIC is also unsuitable for most non biological uses because of its dependence on polarization, which many physical samples would affect.

One non-biological area where DIC is useful is in the analysis of planar silicon semiconductor processing. The thin (typically 100-1000 nm) films in silicon processing are often mostly transparent to visible light (e.g., silicon dioxide, silicon nitride and polycrystalline silicon), and defects in them or contamination lying on top of them, become more visible. This also enables the determination of whether a feature is a pit in the substrate material or a blob of foreign material on top. Etched crystalline features gain a particularly striking appearance under DIC.

Image quality, when used under suitable conditions, is outstanding in resolution and almost entirely free of artifacts unlike phase contrast. However analysis of DIC images must always take into account the orientation of the Wollaston prisms and the apparent

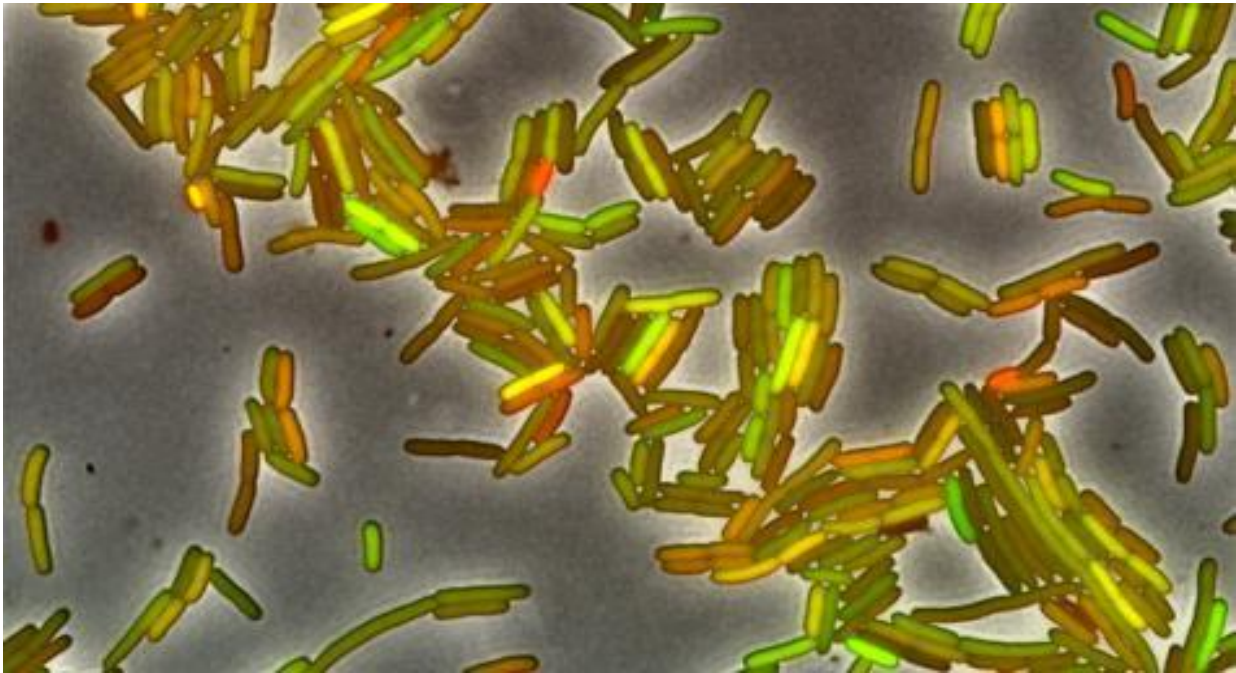


Figure 4: Gene expression is inherently noisy [15]. These cells express two fluorescent proteins, one shown in red, and the other in green. Both genes are controlled by the same promoters; therefore, if gene expression were deterministic, all cells would have equal amounts of red and green (and therefore appear yellow). The presence of cells that are more red than green, or vice versa, results from stochastic fluctuations (noise) in gene expression.

lighting direction, as features parallel to this will not be visible. This however, can be easily overcome by simply rotating the sample and observing changes in the image.

1.2.2 Confocal Laser Scanning Microscopy

Confocal laser scanning microscopy (CLSM or LSCM) is a technique for obtaining high-resolution optical images with depth selectivity. The key feature of confocal microscopy is its ability to acquire in-focus images from selected depths, a process known as optical sectioning. Images are acquired point-by-point and reconstructed with a computer, allowing three-dimensional reconstructions of topologically complex objects. For opaque specimens, this is useful for surface profiling, while for non-opaque specimens, interior structures can be imaged. For interior imaging, the quality of the image is greatly enhanced over simple microscopy because image information from multiple depths in the specimen is not superimposed. A conventional microscope "sees" as far into the specimen as the light can penetrate with decreasing contrast quality, while a confocal microscope only "sees" images one depth level at a time. In effect, the CLSM achieves a controlled and highly limited depth of focus. The principle of confocal microscopy was originally patented by Marvin Minsky in 1957, but it took another thirty years and the development of lasers for CLSM to become a standard technique toward the end of the 1980s. In 1978, Thomas and Christoph Cremer designed a laser scanning process, which scans the three dimensional surface of an object point-by-point by means of a focused laser beam, and creates the over-all picture by electronic means similar to those used in scanning electron microscopes. This CLSM design combined the laser scanning method with the 3D detection of biological objects labeled with fluorescent markers for the first time. During the next decade, confocal fluorescence microscopy

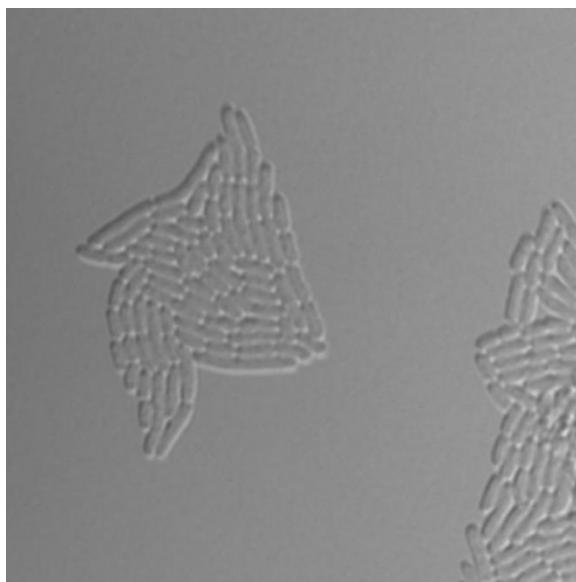


Figure 5: S. Typhimurium micro-colony monitored by DIC microscope.

was developed into a fully mature technology, in particular by groups working at the University of Amsterdam and the European Molecular Biology Laboratory (EMBL) in Heidelberg and their industry partners.

In a confocal laser scanning microscope, a laser beam passes through a light source aperture and then is focused by an objective lens into a small (ideally diffraction limited) focal volume within or on the surface of a specimen. In biological applications especially, the specimen may be fluorescent. Scattered and reflected laser light as well as any fluorescent light from the illuminated spot is then re-collected by the objective lens. A beam splitter separates off some portion of the light into the detection apparatus, which in fluorescence confocal microscopy will also have a filter that selectively passes the fluorescent wavelengths while blocking the original excitation wavelength. After passing a pinhole, the light intensity is detected by a photodetection device (usually a photomultiplier tube (PMT) or avalanche photodiode), transforming the light signal into an electrical one that is recorded by a computer.

The detector aperture obstructs the light that is not coming from the focal point, as shown by the dotted gray line in the image. The out-of-focus light is suppressed: most of the returning light is blocked by the pinhole, which results in sharper images than those from conventional fluorescence microscopy techniques and permits one to obtain images of planes at various depths within the sample (sets of such images are also known as z stacks).

The detected light originating from an illuminated volume element within the specimen represents one pixel in the resulting image. As the laser scans over the plane of interest, a whole image is obtained pixel-by-pixel and line-by-line, whereas the brightness of a resulting image pixel corresponds to the relative intensity of detected light. The beam is scanned across the sample in the horizontal plane by using one or more (servo controlled) oscillating mirrors. This scanning method usually has low reaction latency and the scan speed can be varied. Slower scans provide a better signal-to-noise ratio, resulting in better contrast and higher resolution. Information can be collected from different focal planes by raising or lowering the microscope stage or objective lens. The computer can generate a three-dimensional picture of a specimen by assembling a stack of these two-dimensional images from successive focal planes.

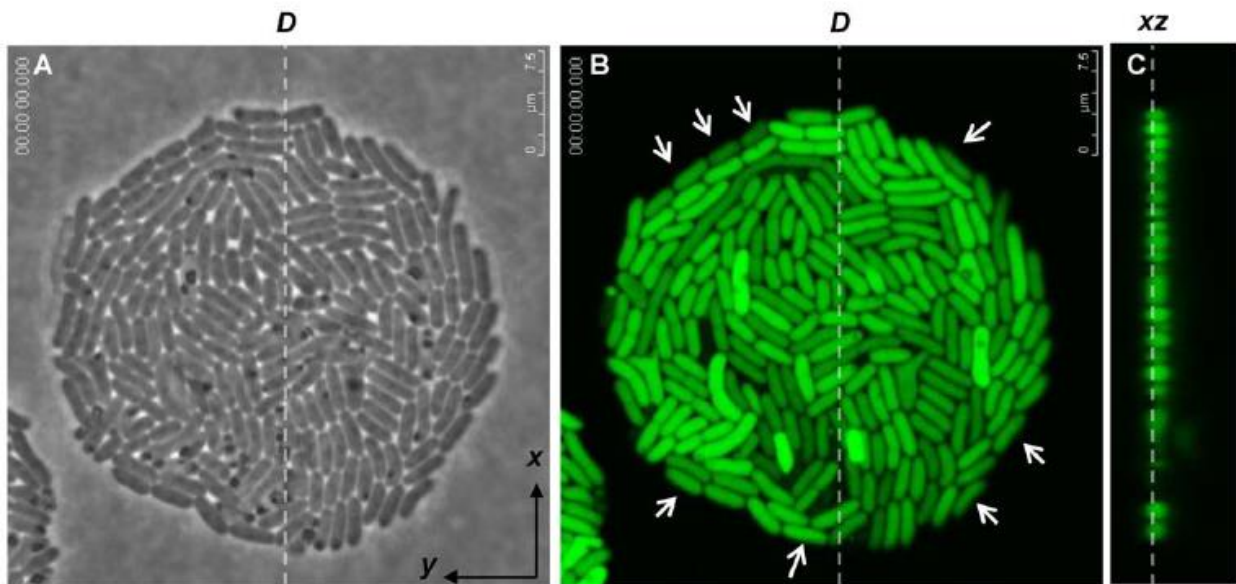


Figure 6: Micro-colony was observed [16]: (A) on the xy plane with phase-contrast microscopy; (B) on the xy plane with CLSM; (C) on the xz plane along the diameter-line “D” with CLSM.

Confocal microscopy provides the capacity for direct, noninvasive, serial optical sectioning of intact, thick, living specimens with a minimum of sample preparation as well as a marginal improvement in lateral resolution. Biological samples are often treated with fluorescent dyes to make selected objects visible. However, the actual dye concentration can be low to minimize the disturbance of biological systems: some instruments can track single fluorescent molecules. Also, transgenic techniques can create organisms that produce their own fluorescent chimeric molecules (such as a fusion of GFP, green fluorescent protein with the protein of interest).

1.3 Necessity for Software Development

The available information on the heterogeneity in the growth behavior of single cells is, in general, limited mainly due to the technical difficulties in monitoring the growth of individual cells. The studies on the variability of individual cell behavior are mainly based on time-lapse microscopy studies and require data from hundreds or even thousands of cells [17-19]. Among the difficulties of the method is the effective analysis of such high number of data. Indeed, monitoring and analyzing the number and the properties of individual cells (i.e. division time, size etc) within a growing micro-colony is a great challenge. Even the estimation of cell counts is currently very time-consuming and prone to errors, since it is essentially performed by the human-user. The commercially available software tools quite often fail completely, or provide very imprecise cell counts. Estimating accurately the number of cells by processing sequences of images of microbial communities and assessing their morphology in a fully automated manner is an important open problem in bio-image analysis. So, new developments in image analysis of microbial growth at a level of individual cell will create new insight and provide effective databases for stochastic modeling approaches and microbial risk assessments [20]. Finally, assessing this open issue will benefit the system biology related community since a high throughput method is essential for microbial community modeling.

1.4 Thesis Outline

The rest of the thesis is organized as follows. In State of the Art chapter we discuss other software packages capabilities and introduce the proposed methodology. In the Methods Chapter we present and describe the developed pipeline. Chapter IV entitled as Datasets and Evaluation Scheme presents the datasets used for evaluation and the evaluation scheme followed. In Chapter V, the Results and Discussion Chapter, a thorough evaluation of the software is presented along with an illustration of several additional features provided by the software. Finally, in the Conclusions and Future Work Chapter, we briefly state some final conclusion and point to future advances and improvements.

2. STATE OF THE ART

In this chapter we will present the state of the art software packages. On the first section, we will describe the existing software packages and discuss their limitations. In the second section, we will summarize the capabilities of the proposed pipeline.

2.1 Existing Software Packages

There are currently multiple image analysis programs capable of cell detection, although most of them work best with eukaryotic cells. Bacterial cells are comparatively small in size, often close to the resolution of optical microscopy, making it challenging to use standard pixel based techniques to separate clusters of cells and to obtain the required subpixel precision. A few image analysis tools have been used for bacterial cells, which typically start with image thresholding in order to outline bright or dim cells against a uniform background, produced using either phase contrast microscopy or fluorescence microscopy of uniformly labeled cells [21-22]. These programs can be categorized into two groups based on their approaches. One focuses on maximizing resolution of cell outlines by utilizing interpolation between pixel values [21]. This method produces high precision cell contours in well-separated cells but fails to identify touching or hard-to-resolve cells, preventing analysis of densely packed cells in still images as well as progeny in time-lapse images of dividing cells. The other group of methods used for bacterial cells images focuses on segmenting the image to separate densely packed cells at the expense of precision by using pixel-based operations, such as edge detection, watershed, and morphological erosion and dilation [22-23]. However, there is a strong demand for the capability of obtaining subpixel precision outlines of touching cells in noisy images while increasing automation of the process.

Several software packages are currently available, where the most widely used and efficient are TLM-Tracker [24], CellTracer [23] and MicrobeTracker [25]. TLM-Tracker employs multiple alternative algorithms for segmentation, namely threshold-based algorithms, watershed transformation and level-set techniques. TLM-Tracker provides the ability of lineage construction, where the cells are tracked by searching of overlapping cell areas in the proceeding frames of the movie. In the CellTracer, the authors developed the concept of hybrid grey-scale/black-white images and they extended existing image filters and mathematical morphological operators for grey-scale images to work with these hybrid images. This approach allows extracting cells from an image iteratively in order to gradually convert the grey-scale image into a binary mask of segmented cells, without relying on initial cell markers. As far as the tracking (lineage computation) is concerned, they incorporated neighboring cell information to compute numerical likelihood scores for cell identity between each pair of time points t to $t+1$ and they applied an integer programming method to generate frame to frame correspondences between cells and the lineage map. MicrobeTracker exploits for the segmentation step several algorithms developed for medical image segmentation and computer vision, which include clustering, template-matching, active contours, region growing and level set methods.

All the aforementioned software packages exhibit several limitations in efficiency that prevents them to perform reliable in a fully automated way. Specifically, the CellTracer requests from the user to have image processing background in order to select the appropriate pipeline for the images at hand and this is repeated even for images of the same dataset throughout the analysis. The same observations hold for the TLM-Tracker. MicrobeTracker was found to be extremely non user friendly due to its complicated parameterization. The developers give some default parameters settings from specific image modalities and organism types, but due to their sensitivity further

Table 1: Software Packages Overview

Software	Tracking Multiple colonies	Robustness to dataset quality	Parameterization	Optical/Confocal images	Phase contrast/fluorescent images
CellTracer	no	no	Complex	Optical	both
MicrobeTracker	no	no	Complex	Both	both
TLM Tracker	no	no	Complex	Optical	both
Developed Methodology	yes	yes	Simple	Both	both

fine tuning is needed. Furthermore, a very significant disadvantage of all the state of the art methods is the lack of generality (overfitting on specific type of data), i.e. they are not able to perform robustly for a wide variety of images acquired by several imaging settings (microscope type, imaging type etc). As a result of those limitations the estimation of cell count and their corresponding features (e.g. length, area, gfp quantification, etc), is currently very time-consuming and error prone. So, it is clear that efficient and high throughput estimation of cells from sequences of images of microbial communities and assessing their morphology is an important and open problem in bio-image analysis. All the aforementioned limitations (Table 1) clearly state that bacterial image processing is still considered as a bottleneck towards a high throughput analysis and the need for efficient and automatic tools (high throughput) is apparent.

2.2 Developed Methodology

In this work, we present a new bacterial cell analysis software, which make use of image processing and machine learning methods, utilizes the specific cell shape and time-lapse information when available, in order to achieve high precision cell detection even in densely packed and noisy images. The presented methodology allows automated outlining and subsequent analysis of cells in both single and time-lapse image sequences. It is governed by the “divide and conquer” principle combining different approaches (from mathematical morphology approaches to machine learning techniques) according to the processing level and image quality. The basic framework is based on dividing the problem to smaller sub-problems. So, we step from image’s level to colony’s level. This enables us to analyze input images independently of the number of micro-colonies they contain. This is a very important novel feature that previous approaches do not offer. So, a scientist may monitor several micro-colonies at the same time and extract their individual feature without the need of cropping and further manipulating the raw image. This way one can save time (experimental time for data gathering) and money (several assays can be analyzed simultaneously).

In colony’s level, we divide again the problem in order to successively result to single cells. This approach gives us the capability of analyzing colonies regardless of their population, i.e. over-crowded images. An interesting property of our methodology is that once we reach the colony’s level, regardless of the image origin, the pipeline operates

using the same algorithmic pipeline. Also, this property is unique among other software packages and it is the one that provide its generality property. The pipeline is fully automated, i.e. high throughput, and does not need human interference even for different input image modalities. Finally, the robustness of the segmentation enables the proposed software to extract valuable information from the input dataset which can be visualized according to the needs of human user.

3. METHODS

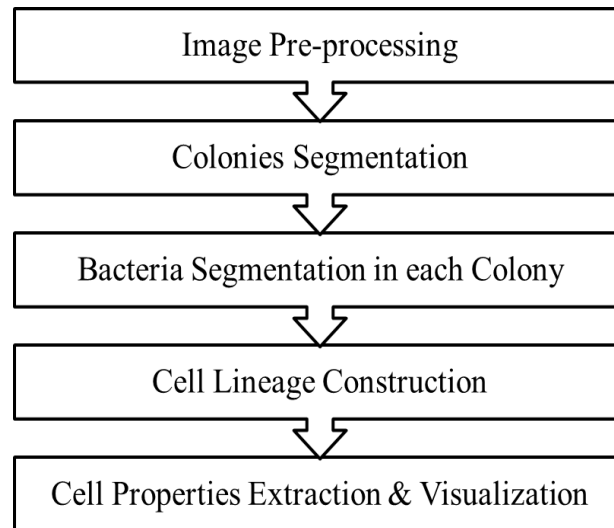


Figure 7: Proposed pipeline overview

In this chapter we will present in detail the developed pipeline, which consists of five (5) consecutive steps, i.e. preprocessing, colonies segmentation, bacteria segmentation, cell lineage construction, features extraction and visualization. The whole pipeline is designated by a continuous cascade process in order to detect the individual single cells. The pipeline starts from an image, segments the colonies, detects objects with one and/or more cells and finally gets to the desired result, the single cell detection. Next we will describe in detail the developed pipeline and illustrate that the aforementioned “journey” from image to cell is not only necessary but also gives valuable information along the way. The pipeline stages are summarized in Fig. 7.

3.1 Preprocessing

The pipeline starts with image preprocessing, a very common and necessary process used for noise suppression and background uniformities elimination. Also, in our methodology, the preprocessing embeds a first level of segmentation, in terms of colony segmentation. In more detail, we apply the image denoising methods described in [26-28], in order to suppress the inherent noise. Then, we apply contrast-limited adaptive histogram equalization (CLAHE) as described in [29] so as to emphasize only the bacterium related region instead of the noisy and luminous background situated locally at a colony area. The effects of these actions are illustrated in Fig. 8. CLAHE operates on small regions in the image called tiles rather than the entire image. Its tile’s contrast is enhanced so that the histogram of the output region approximately matches the histogram specified by the uniform distribution. The neighboring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas can be limited to avoid amplifying any noise that might be present in the image. Consequently, the noise of the initial image is removed and the pixel belonging to cell areas are separated from colony’s background.

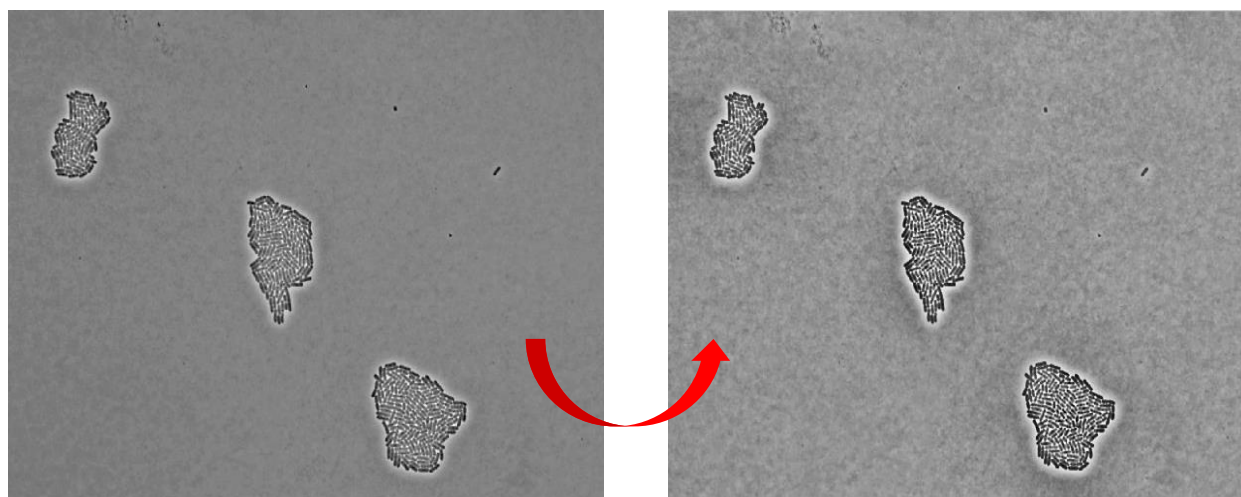


Figure 8: Input image in preprocessing stage (left). Output image (right) after denoising and CLAHE application.

3.2 Colonies Segmentation

Following that, we developed a colony segmentation method, where a binary mask segmenting the bacterial colonies from the image background is created. For this step, we first employ mathematical morphology operators [30] and then apply the Otsu's global thresholding algorithm [31] so as to create the binary image, outlining as precisely as possible the area of each colony. However, at this stage, the resulting binary image contains small artifacts (due to the image background) additionally to the rigid colonies. In order to cope with this, an edge detection algorithm, Canny's algorithm [32] is used to the preprocessed image. The output of Canny's edge detector produces a binary image containing only the edges of each colony. Then, by applying morphological dilation to the edge detected image, we construct a binary image containing a rough mask of each colony relieved from noisy artifacts. Finally, using a simple multiplication operation of the mask created by global thresholding and the mask created by edge detection, we get a mask outlining in detail each colony boundary (see Fig. 9 for details).

As a result, we come up with an image where each colony (regardless of its size) is segmented and outlined. As mentioned, and it will be more obvious at the next sections, the proposed method is based on "divide and conquer" principle. So, from this step and forward, each colony will be processed separately and save its properties separately (i.e. colony tracking through time, given a time series experiment). This feature is very significant since it is very common for a colony (in a multi-colony experiment) to: 1) merge with a neighbor colony, 2) move out of the microscope's field of view, or 3) continue to grow individually. So, keeping track of colonies is vital for archiving their corresponding properties (number of bacteria, etc.) through time.

3.3 Single cell Segmentation Algorithm

In this section we will look into the heart of the developed methodology, which final result is single cell segmentation. Having already defined the bacterial colony areas as accurately as possible, we need to "zoom in" and detect, in an efficiently way, the individual cells.

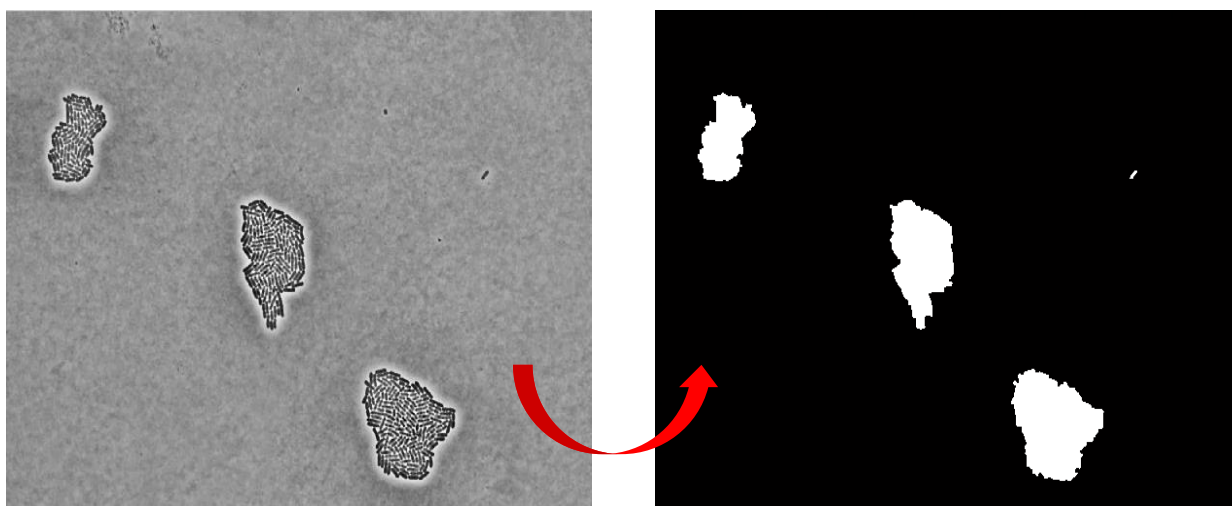


Figure 9: Input image in colonies segmentation stage (left). Output image (right), a mask with segmented colonies.

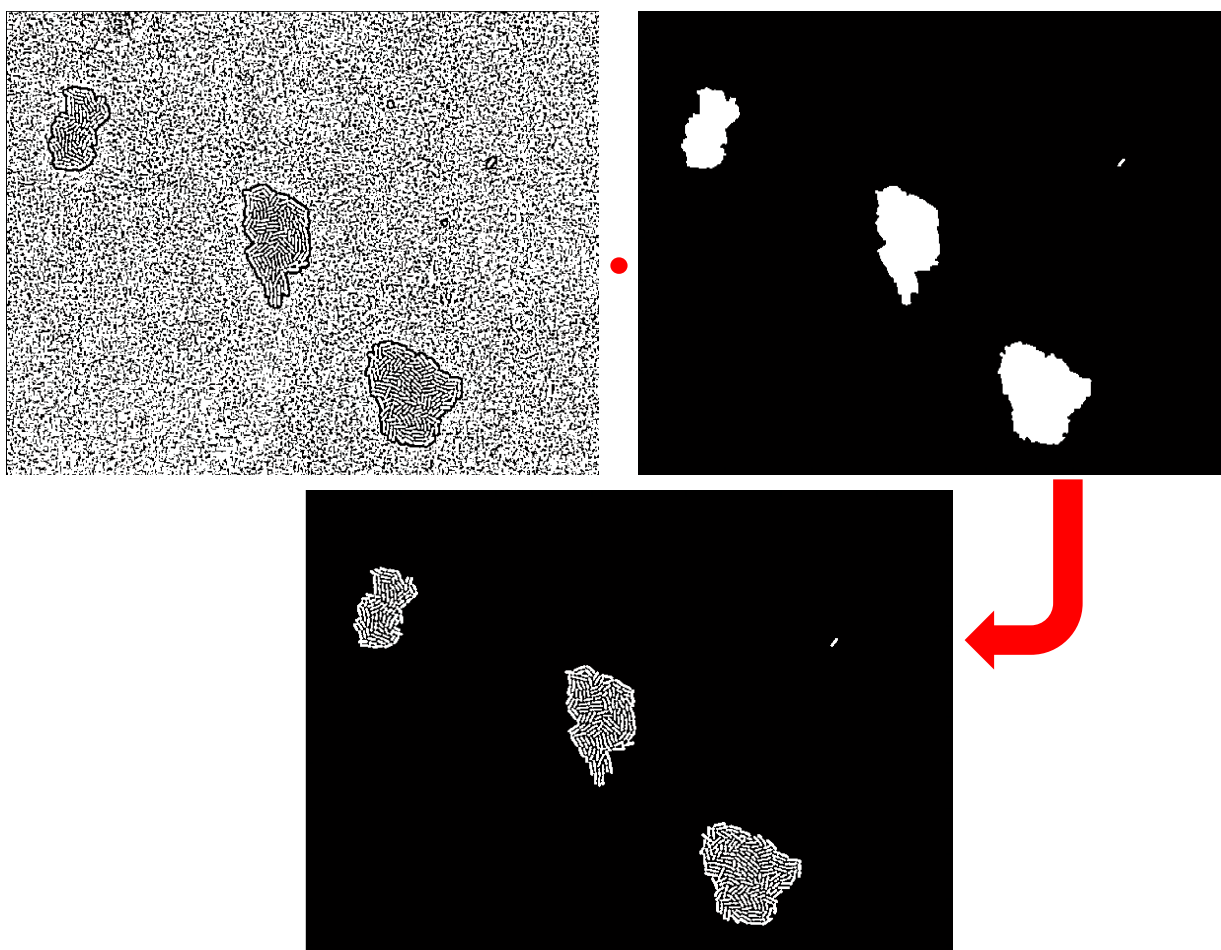


Figure 10: Initial segmentation image (bottom) created by multiplication of the two masks (top)

Due the large illumination variations in the images (locally and globally), the previously identified as foreground areas, include not only cell areas but also a varying degree of local background, we apply an adaptive thresholding algorithm [30] to the preprocessed image. Thus, we manage to separate the foreground (bacteria) from the non-uniformly

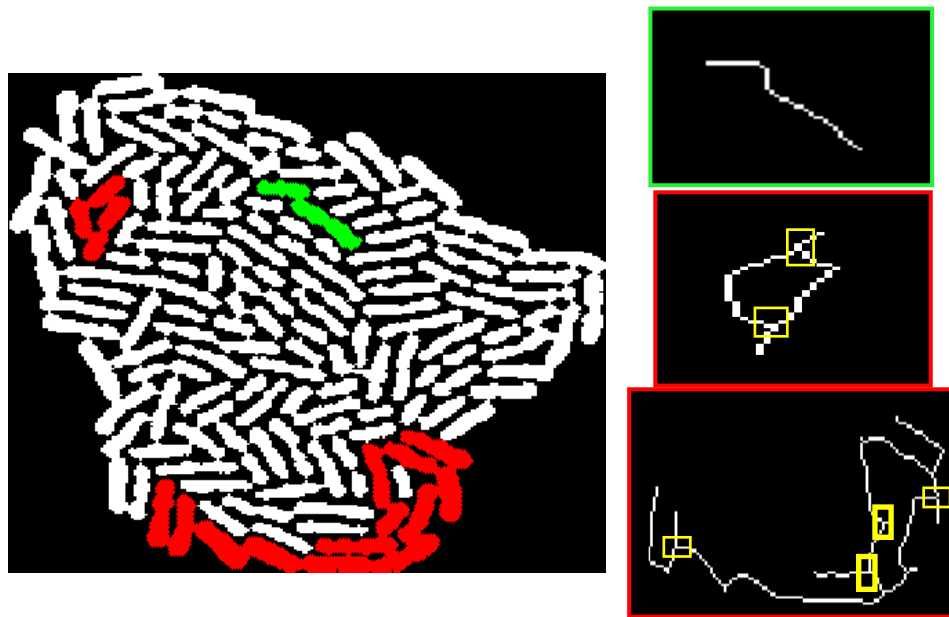


Figure 11: Skeletonization of three objects. The object of the top panel (green) has no junctions, i.e. collinear object, while the objects in the rest of panels (red) have more than one junction, i.e. complex object.

illuminated background (please refer to Fig. 10 top left) but we create small artifacts to the background image regions, which are eliminated by multiplying this image with the mask created in the previous step. We must note that this type of background (in the foreground's objects) is very common in images acquired by optical microscopes (either bright field or phase contrast) and more limited by confocal microscopes. Additionally the adaptive threshold used, has no negative effect on confocal images, i.e. it performs equally well (in terms of excluding local background pixels) in both image types. As it is observed in the Fig. 10 bottom while the most of the local background is removed, the result is not ideal, i.e. each segmented object does not correspond to a single bacterium. For this, we developed a novel methodology where we treat each object individually.

First, for each object, its skeleton is computed [30] (refer to Fig. 11). Then, if the skeleton has junctions, the object is considered as a complex object, i.e. is a cluster of several single bacterial cells. We will describe later the process when a complex object is detected. In the case that no junctions are found, the object may be either a distinct bacterium or two collinear bacteria (collinear object). In the former case the object must not be further segmented (single bacterium) and the segmentation process stops while in the latter the object must be segmented.

3.3.1 Collinear Object Processing

We first review the simpler case of no junctions found at the object's skeleton. In order to classify automatically an object into these two categories we developed a "smart" criterion, called "deep" valley criterion. Its functionality relies on the difference of the physical shape of an object when representing a bacterium at the fission stage [33] (similar to a bow tie) or not (please refer to Fig. 12).

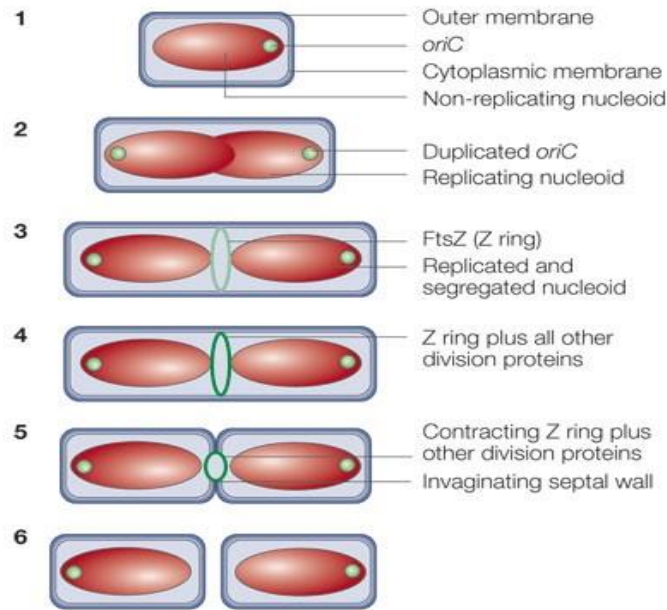


Figure 12: Bacterial fission [33]. In stage 5 the cell enters into the division stage. Our method detects objects having the shape of stage 5, i.e. bow tie shape.

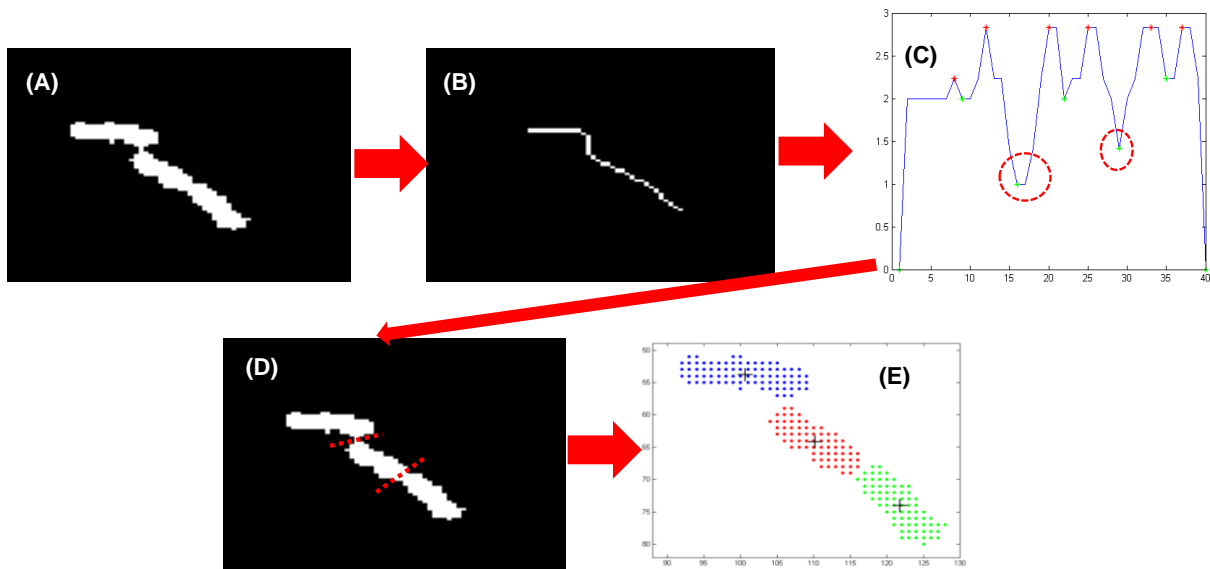


Figure 13: Pipeline of identification of bowties in collinear objects

In order to “quantify” the shape of the object and use it for classification we compute the Euclidean distances [34] of the object along its centerline and search for a local minimum (“deep valley”) (Fig. 13 panel (C)). If such a local minimum exists and fulfills the following ratio constraints, then is considered to be a “deep valley” and T is a threshold which depends on input image resolution, i.e. how many pixels form a bacterium. We must note that the value of T is automatically set by the software when the resolution is inserted by the user and it is in the range of $[0.65, 0.85]$ for low to high resolution. When the criterion classifies the object as a bacterium being in the division phase, the object is split into two discrete bacteria exactly where the “deep valley” is found (Fig. 13 panel (C) red circle), otherwise the segmented cell area remains intact.

3.3.2 Gaussian Mixture Modeling

In this section we begin with a brief presentation of GMMs, a powerful pattern recognition tool that is used extensively in the Expectation-Maximization framework [35]. For more details the interested reader is referred to [36]. Mixture models, is an extremely useful tool which provides the ability of model-based approach to unsupervised clustering. Complex multimodal probability density functions (PDFs) can be represented simply using the models of the Gaussian mixture. This is why they can also be used as a method for elaborate class-conditional probability density functions representation in supervised learning scenarios. The application of mixture modeling here amounts to fitting Gaussian mixtures to the 2D-observations without the knowledge of the optimal number of components needed. Let $\mathbf{Y} = [Y_1, \dots, Y_d]^T$ be a random variable of dimensionality d , with $\mathbf{y} = [y_1, \dots, y_d]^T$ representing a specific realization of \mathbf{Y} . Then \mathbf{Y} has a finite mixture distribution if its probability density function can be written as:

$$f_Y(\mathbf{y}|\boldsymbol{\theta}) = \sum_{m=1}^C \pi_m f_Y(\mathbf{y}|\theta_m) \quad (1)$$

where $f_Y(\mathbf{y}|\boldsymbol{\theta})$ is a component density function, C is the number of components and π_m are their mixing probabilities. Having a specific density function, common for all the mixture components yields a parameter set $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_C, \pi_1, \dots, \pi_{C-1}\}$ where it holds that $\pi_C = 1 - \sum_{m=1}^{C-1} \pi_m$. For the developed methodology, the Normal (Gaussian) density function was considered, thus $f_Y(\mathbf{y}|\theta_m) = N(\mathbf{y}|\boldsymbol{\mu}_m, \Sigma_m)$ with a general covariance matrix Σ_m and mean vector $\boldsymbol{\mu}_m$. Hence, in that case the component parameters are $\theta_m = (\boldsymbol{\mu}_m, \Sigma_m)$. The maximum likelihood (ML) estimate of the mixture parameters $\boldsymbol{\theta}$, based on a set of n independent observations $\mathbf{y} = \{\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n)}\}$, is

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mathbf{y}) \quad (2)$$

where $L(\boldsymbol{\theta})$ is the likelihood of the data set under the model. As it is known, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ does not have in general a closed form expression but may be approximated iteratively by applying the EM algorithm.

3.3.3 Complex Object Processing

It is obvious now that when we have a complex object to analyze, it is much more complicated and another approach is employed. In that case, we initially estimate the number of candidate (possible) bacterial cells residing in each object, using a combination of the 1-nearest neighbor (1-NN) algorithm [37] and the watershed algorithm [38]. To do so we create a distance matrix of the binary object using the Chessboard distance transform. For each object's pixel, the distance transform computes the distance between that pixel and the nearest non-zero pixel of the complementary binary image. We then take its complement, force non-object pixels to be *-Infinite* and compute the watershed transform on that. In order to exclude some regions of the object's background erroneously included by the watershed transform we multiply the watershed label matrix with the object's binary mask (see preprocessing Section). At this stage some pixels labeled as 0 do not belong to a unique watershed region and classified to a watershed region according to the nearest neighbor rule. Finally, we determine the centroid of each watershed region, and they represent the initial centers of the cells into the complex object (Fig. 14 panel (4a)).

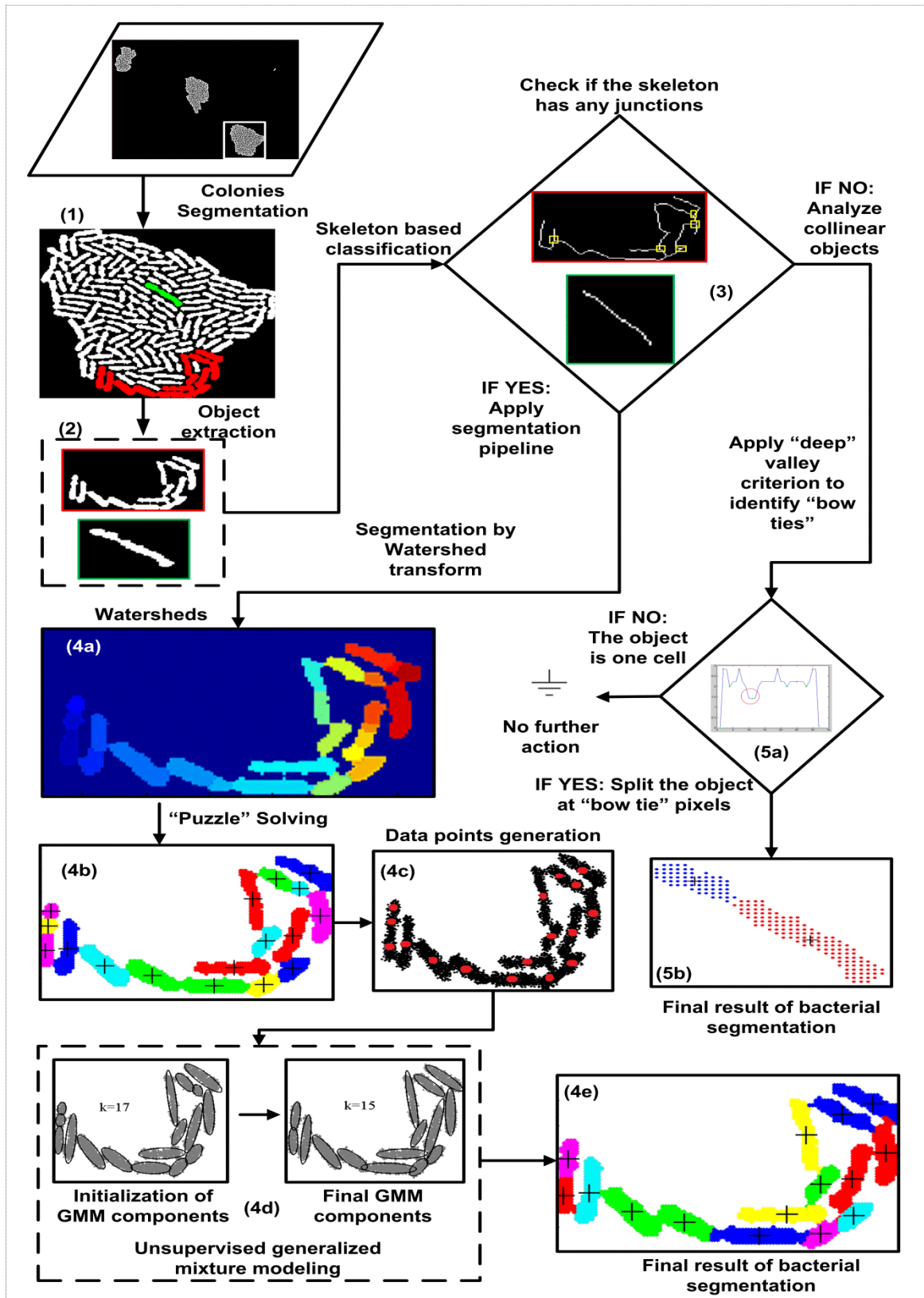


Figure 14: Schematic overview of the proposed segmentation methodology. (1) Colony segmentation, (2) object extraction, (3) skeleton based classification; (4) Complex object segmentation: 4a) Segmentation by watershed algorithm, 4b) "puzzle solving" step, 4c) dataset generation, 4d) unsupervised mixture modeling, 4e) final result of bacterial segmentation; (5) Collinear object segmentation: 5a) Application of "deep" valley criterion and identification of "bow ties", 5b) final result of bacterial segmentation. At step (3), there is a bifurcation leading to two different processing routes, according to the object type that is going to be segmented.

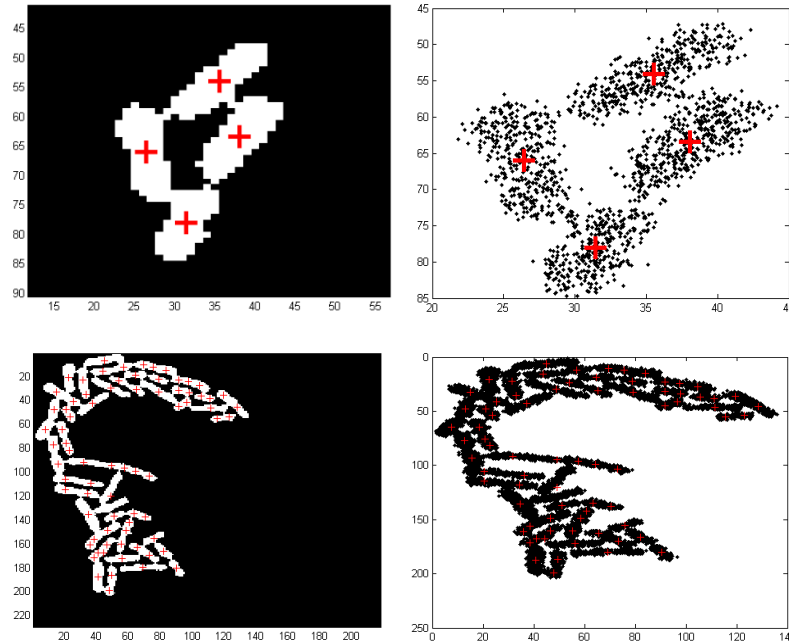


Figure 15: Dataset generation examples. The dataset's size is proportional to the object size.

Another issue that we have to surpass is the inherent property of the watershed transformation, oversegmentation. To do so, we again apply the aforementioned idea of “deep valley” criterion in a slightly modified way. Now, the algorithm attempts to merge the erroneously segmented fragments by following specific “constraints”. All pairs in the neighborhood are exhaustively examined under the criterion if they should be merged or not. The touching objects will be merged only if there is no “bowtie” identified by the criterion. If an object can be merged with more than one neighboring object, then the merging with the maximum solidity is chosen (solidity is defined as the ratio of area to the corresponding convex hull area). Using the prior knowledge of the elliptical shape of the bacteria, we can assume that correct single cell objects tend to have solidity close to one. So the new object (after possible merging) is also inserted to a queue in order to further examine if it should be merged with another watershed fragment. The developed methodology can be regarded as “puzzle” solving, and as we will demonstrate in the Results Chapter, is robust and efficient (Fig. 14 panel (4b)).

As a final refinement of the segmented single cell objects and in order to outline the boundaries of each detected cells as tight as possible we used a machine learning approach based on GMM, described in [39]. In [39], the authors transform the pixel intensities into data points for GMM modeling. In our case, as one can observe from Fig. 16, the intensities of cells are not trustworthy since they do not exhibit a uniform distribution over the cell but a rather noisy one. In order to apply a similar technique, we use the distance matrix as reference for data point generation (Fig. 14 panel (4c) and Fig. 15) instead of the intensities. More specifically, we assume each pixel as a data generator in its neighborhood. The total number of data points N to be generated for each image object will be proportional to the number of the already estimated bacterial centers C . This pool of N data points are distributed among the pixels of the object according to their relative distances (in the sense that a “more internal” pixel will “throw” more points). The physical meaning of this transform is that the pixels closer to the object's centerline are part of it with higher probability than the more distant ones.

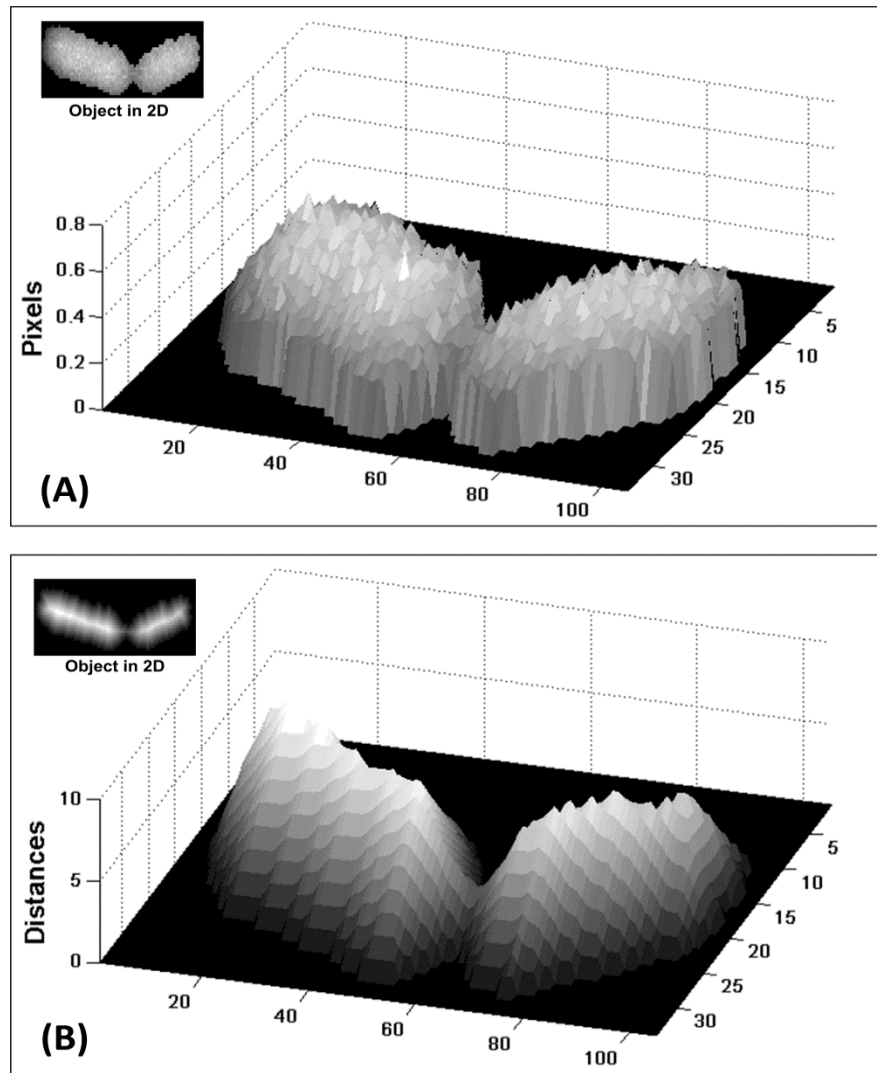


Figure 16 Panel A: Representation of an image object in 3D space, the pixels intensities lie in the z axis. Panel B: Representation of the same object distance transform in 3D space, the distance values lie in the z axis. It is obvious that distance transform smoothes object abnormalities while sharpens the valley between the two bacteria comprising the object.

We may think of this process (moving from pixel distances to a data points distribution) as a reverse engineering process where the resulting data points represent the bacterial elliptical shape whose outline is reflected as the cell pixels on the image. Specifically, a pixel i with distance d_i acts as 2-D Gaussian generator $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}=(x_i, y_i)$. Finally, all the pixels belonging to an object, constitute a GMM [35] having as many components as their number (M). Each component is assumed to have a mixing coefficient proportional to its distance value and equal to

$$\pi(i) = \frac{d_i}{\sum_{j=1}^M d_j}$$

For each pixel i , we draw $N \cdot \pi(i)$ data points, from a 2-D Gaussian distribution centered at the pixel's location and having diagonal covariance matrix $\boldsymbol{\Sigma}$ with both its elements set equal to 0.3. This value (0.3) for the variance was selected to be smaller than 0.5 (the half-distance between neighboring pixels) in order to ensure that data points generated by the model (representing "cell structure") will be distributed in a manner that guarantees that their abundance reflect the distance as presented in Fig. 16 in 3-D. Thus preventing the generation of "hills" of data points in-between pixel locations. This

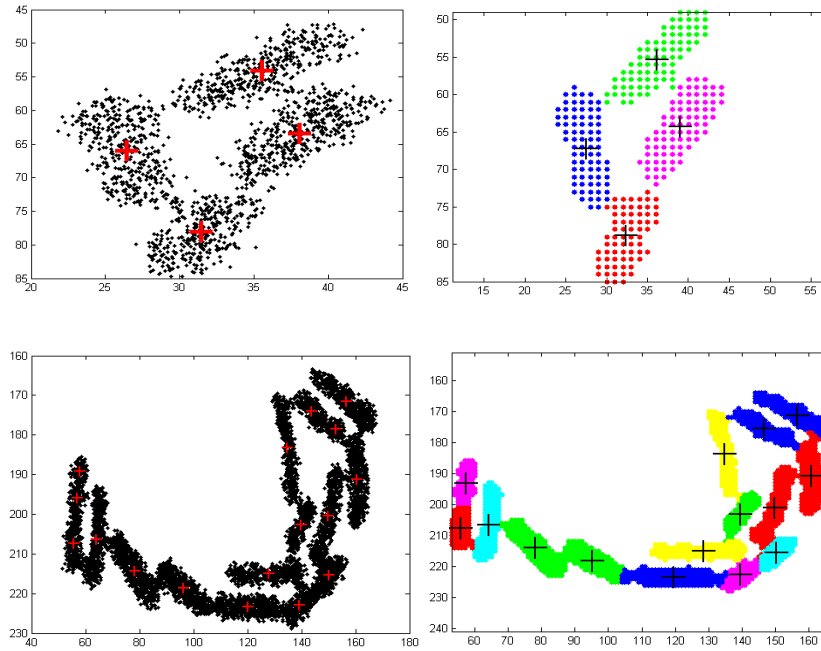


Figure 17: Application of EM algorithm with MML criterion for model selection. Top: No center reduction. Bottom: The centers are reduced from 17 to 15.

value of the variance was determined by experimentation and has been kept fixed throughout the analysis.

As mentioned before, the total number N of generated data points is analogous to the number of the estimated cell centers C and not to the number of pixels in the image object. The justification of doing so is the observation that the area of an object does not determine its complexity i.e. how many different bacteria holds. On the other hand, the number of identified candidate centers can be such an indicator since it approximates the number of cells expected to be present in the image object region. Consequently, if an object has a lot of candidate centers, it is justified to “spend” more data points to capture adequately the underlying distribution of the different bacteria laying in it.

The first step of the 2-D Gaussian modeling process after data generation is initialization, i.e. initial class assignment by computing the log-likelihood of each data point to belong to the initial mixture model, where the complete set of mixture model parameters and w_m are the mixing coefficients.

The last step is the application of Finite Mixture Models (FMM) [35] in two dimensions. Using the so far extracted information on the generated data as initialization we apply a modified EM algorithm [36] which employs the MML criterion [40] for model selection (see Fig. 17). This is done in order to identify the model that fits best to the data points. It is known that EM algorithm tends to overfit the data and favors the more complex solution which in our case translates to a solution with more components than the true cells in the image object. The MML criterion ensures that the final model will not be a complicated one unless it pays for itself. Therefore, the best solution (according to MML) may contain less than the initial C components; i.e. results to an FMM that explains the generated data points better.

However, due to the quality of the image dataset, some cell fragments remain unmerged. So as to avoid this, we prune all the remaining objects by checking whether their circularity is under a pre-specified threshold $Circ$. Circularity is defined as the ratio of minor axis length to major axis length. The bacteria are assumed to have an elliptical

shape, so the oversegmentation pieces tend to have more circular shape so their circularity tend to be close to 1. Furthermore, we reject objects having area (in pixels) under a pre-specified threshold A , which is computed automatically, according to the input image resolution. The rejected objects are finally merged with one of their touching neighbors according to maximum solidity criterion, mentioned previously.

3.4 Cell Lineage Construction

A bacterium in time series can be found in three different states: 1) at the growing state (or stay unchanged), 2) at the division state, and 3) disappear from the microscope's field of view. In order to construct the lineage of a single cell, there are two fundamental requirements. The first is an efficient segmentation of the bacteria of each colony in time series and the second is to track the bacteria along the consecutive frames.

Having two consecutive frames, the contours of the first frame's bacteria (frame $n-1$) and the centroids of the second frame's bacteria (frame n) are extracted, (refer to Fig. 18 (2) for details). Then, the algorithm matches the centroids of the n th frame with the contours of the $(n-1)$ th frame (Fig. 18 (3)). To achieve this, we check whether the bacterial centroids of frame n lie into the bacterial contours of frame $n-1$ or not. If so the bacterium is matched with its "ancestor", otherwise, we continue to search for a match among other cells in the neighborhood. The neighborhood is defined as the bacteria lying inside a circle, with its center coinciding with the given centroid and its radius equals to twice the bacterium's length. So defining the unmatched bacterium's neighborhood, we apply nearest neighbor algorithm to match the centroid to one of the neighboring bacteria, giving priority to bacteria which have not already found their "descendants". The nearest neighbor rule is modified so as to assign a centroid to the nearest bacterial contour, by computing the Euclidean distances between the centroid and the bacterial contours' pixels. These steps are repeated for all the consecutive frames of time series. Finally, the extracted tracking information is used to create a binary tree, which represents the lineage of a starting single cell (at time zero) (e.g. Fig. 34 and 35).

Fig. 19(1) presents the correct pattern of lineage tree structure. By definition, the lineage trees must have all their leaves at the same depth (complete binary tree). Consequently, the patterns presented at Fig. 19 (2) and (3) are erroneous. The first erroneous pattern, which can be called false positive pattern, appeared due to the detection of a false positive bacterium at the segmentation process. In order to correct this error and make the binary tree complete, we merge the erroneous node (false positive, red square) with its sibling. The case of the second erroneous pattern, the false negative pattern happens due to the false detection of a nonexistent bacterium-artifact. To make the binary tree complete, the algorithm splits the putative sibling of the missing node to form two discrete bacteria. We must note that the algorithm, in order to split an object, exploits the segmentation pipeline concerning a complex object (see Section 3.3.3) having the constraint that two discrete objects must be left at the end of this operation.

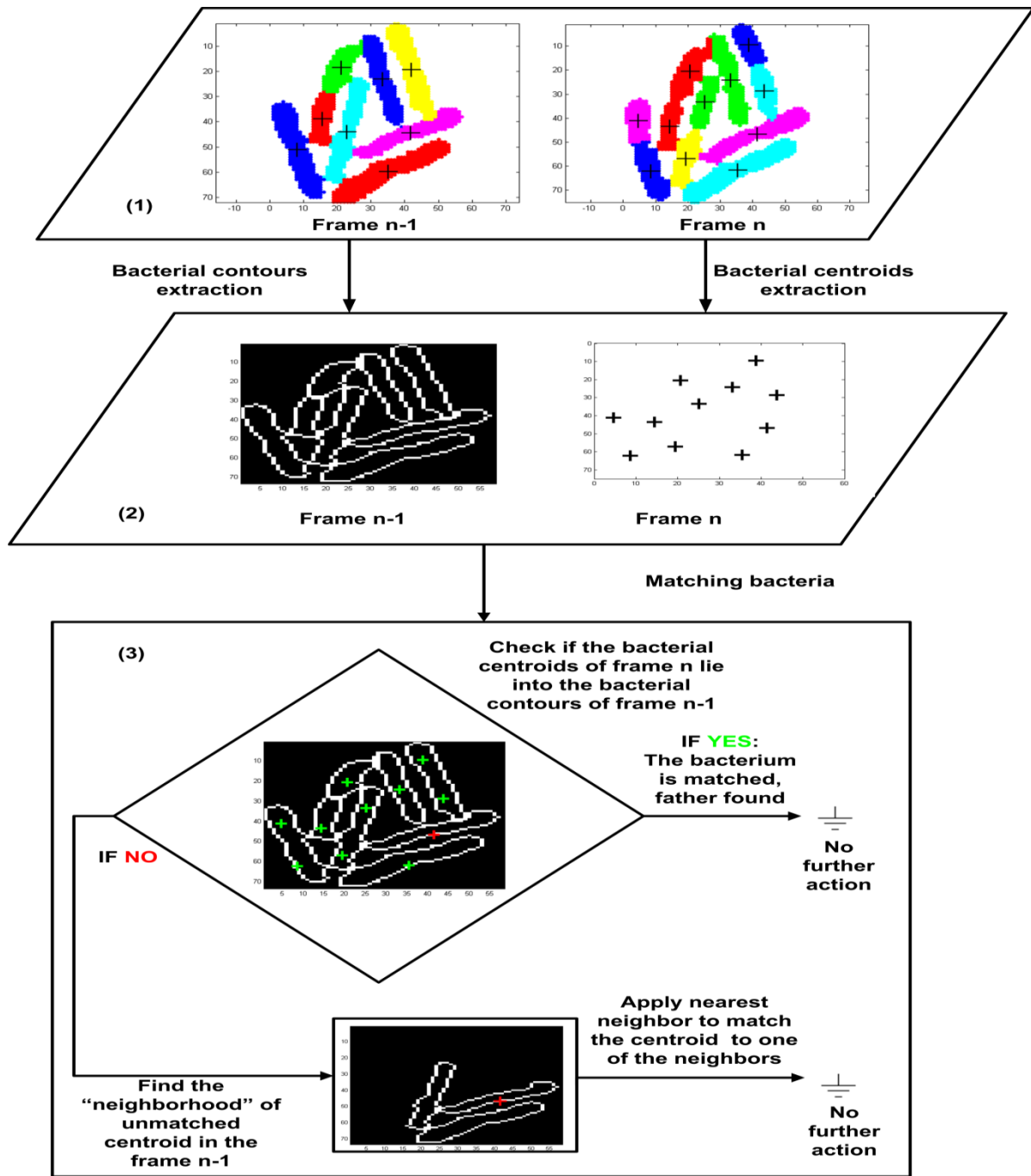


Figure 18: Lineage construction algorithm overview. (1) Segmented bacteria of two consecutive frames, (2) Bacterial contours and centroids extraction from frames n-1 and n respectively, and(3) Bacteria matching.

One significant property that can be derived from the proposed cell lineage construction algorithm is that it can also contribute to the segmentation result refinement. This capability arises from the problem definition, and consequently the lineage tree definition. Other datasets in which bacterial death occurs do not permit those refinements because the lineage trees are allowed to be incomplete.

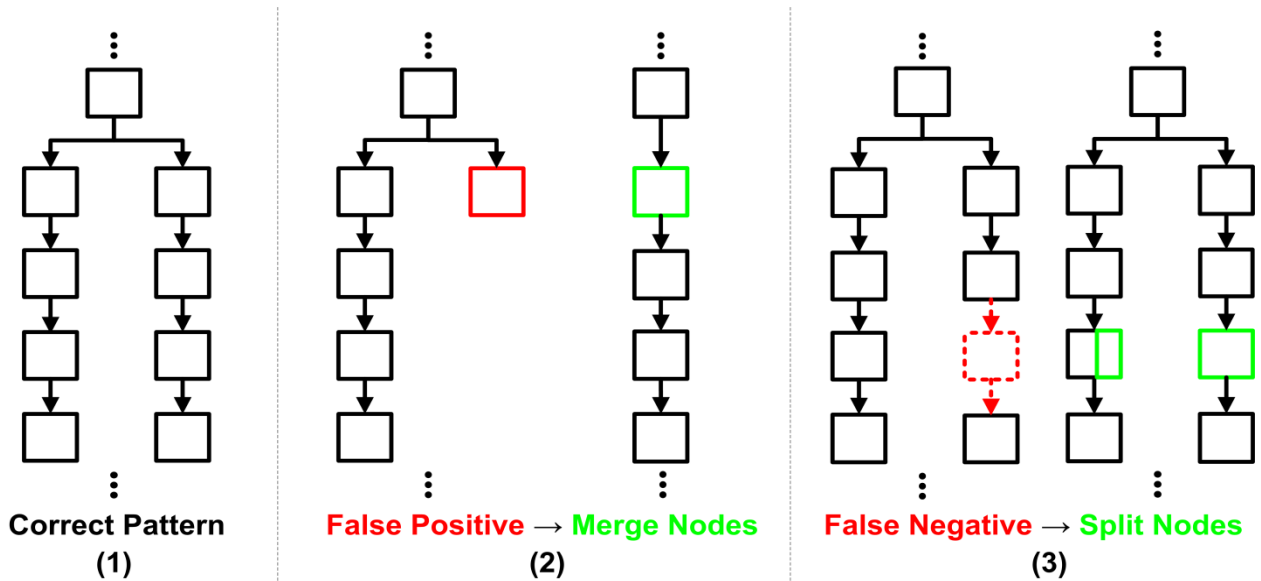


Figure 19: Possible lineage sub-tree patterns. (1) Error free pattern, (2) False positive pattern (red rectangle) and correction (green rectangle) by merging, and (3) False negative pattern (red rectangle) and correction (green rectangles) by splitting.

3.5 Feature Extraction & Visualization

After bacteria segmentation and cell lineage construction, the proposed pipeline extracts several single cell and colony based features. Specifically, the single cell features that can be extracted are: 1) area, 2) perimeter, 3) major axis length (length), 4) minor axis length (width), 5) relative position in the colony, 6) bacterial fluorescent protein quantity, and 7) division time between bacterial generations. All those features can be measured and exported either for each bacterium individually or for each bacterial generation (only in time series). As far as the colony level is concerned, the method computes colony growing rates in terms of area (either in pixels or in micrometers) and the number of bacteria (solution to the bacterial counting problem).

Another important feature of the developed methodology is the ability to produce several naive, yet useful, visualizations. Having extracted all the aforementioned rich information it is vital for a scientist to be able to “see” it. The method can create visualizations either by overlaying the result on input time series, e.g. bacterial area visualization (Fig. 29 line (1) middle), or by creating feature analysis graphs, e.g. histogram of bacterial length (Fig. 29 line (1) rightmost). Further visualizations include the cell lineage of all the aforementioned single cell features, e.g. bacterial area visualization (Fig. 35). Finally, the user can perform statistical analysis by exporting the extracted data to excel files, while they can visualize single cell’s features in order to observe their variations with the naked eye.

4. DATASETS & EVALUATION SCHEME

In this chapter, we will describe the datasets used in order to evaluate the proposed pipeline. Additionally, we discuss the evaluation scheme followed.

4.1 Datasets' Description

Throughout the development and evaluation of the proposed pipeline we used several datasets from different labs and image modalities, in order to ensure that the algorithm is as generic as possible. Here, due to the lack of space, we will present the evaluation results from two datasets that exhibits the most of the capabilities and advantages of the developed software over the current state of the art. The first time series dataset starts with four single salmonella cells(Fig. 20), named from now on as salPhase dataset and acquired by phase contrast optical microscopy(please refer to [41] for further details). It consists of 86 consecutive frames (5 minute period) where the initial growing stages of four micro-colonies are monitored.

The second time series dataset starts with a micro-colony of four bacteria acquired with DIC confocal microscopy. The bacteria were genetically modified by *Tampakaki et al.* to induce green fluorescent protein (GFP) as follows: *Escherichia coli* and *Salmonella typhimurium* ST474 strains were grown in Luria-Bertani (LB) medium at 37°C. The plasmid pDSK-GFPuv [42] was generously provided by Dr. K. Mysore (Samuel Roberts Noble Foundation, Inc.) and was transferred from *E. coli* to *S. typhimurium* ST474 by electroporation (GenePulser, Bio-Rad) following the manufacturer's instructions. Transformants, containing the plasmid, were selected on LB agar plates supplemented with kanamycin at the concentration of 50 µg/ml. We will refer to it as confocal GFP dataset throughout the text (Fig. 21). Specifically, it consists of 43 consecutive frames (5 minute period) where the growing stages of the micro-colony are monitored.

In order to provide an objective and solid evaluation, we additionally test the proposed pipeline with images from several other laboratories. Specifically, our software was evaluated using single frames already used by current state of the art software packages. The first, used by MicrobeTracker, is a phase contrast confocal microscopy image of sparse *E.coli* bacteria and small micro-colonies (please refer to Fig. 22 for details).

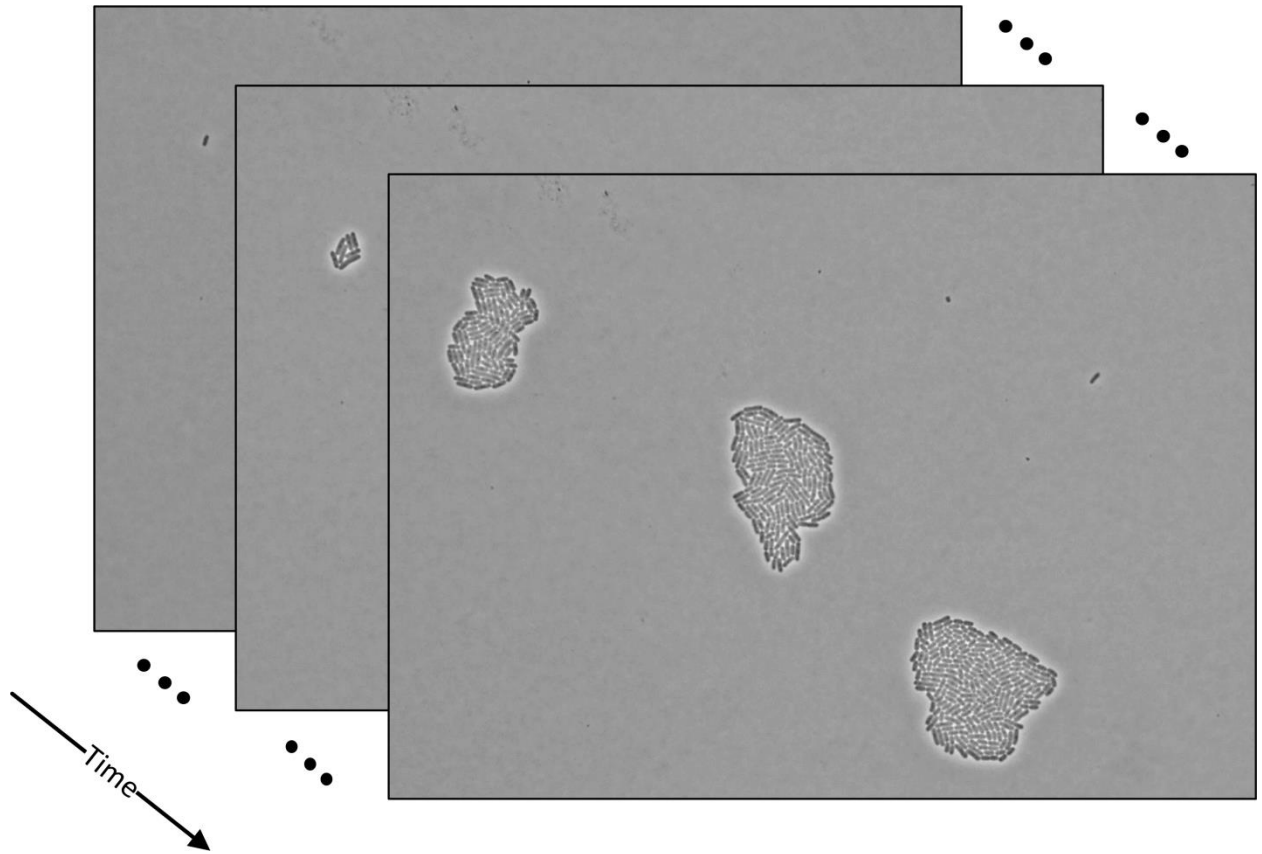


Figure 20: Images' stack of SalPhase dataset.

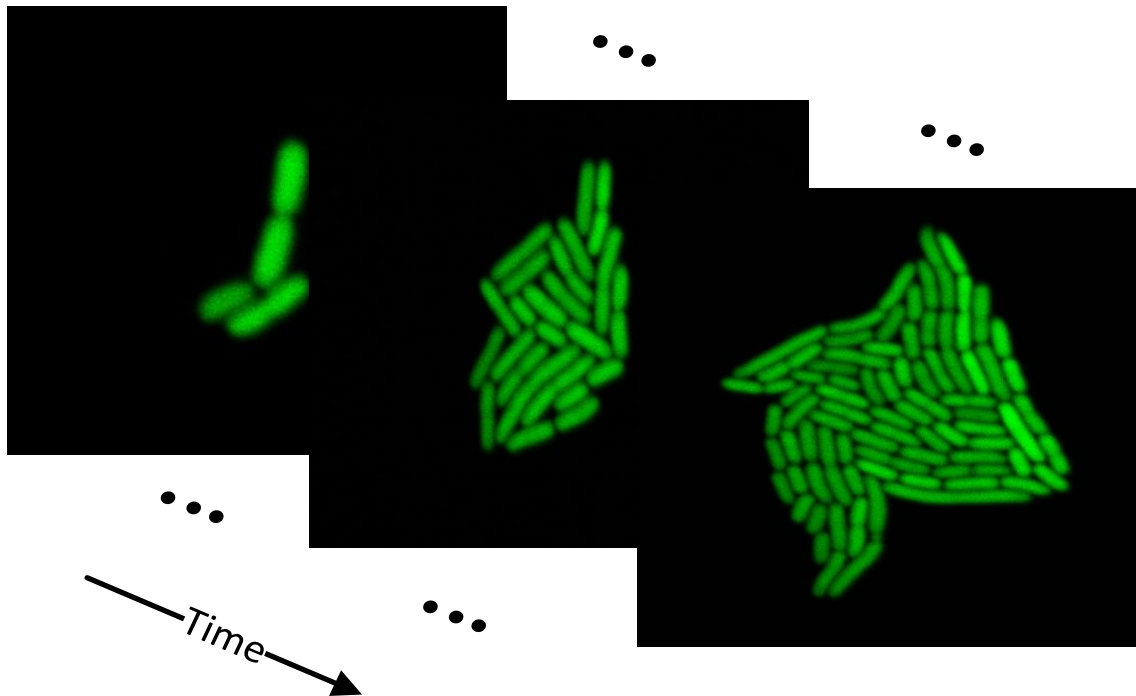


Figure 21: Images' stack of GFP confocal dataset.

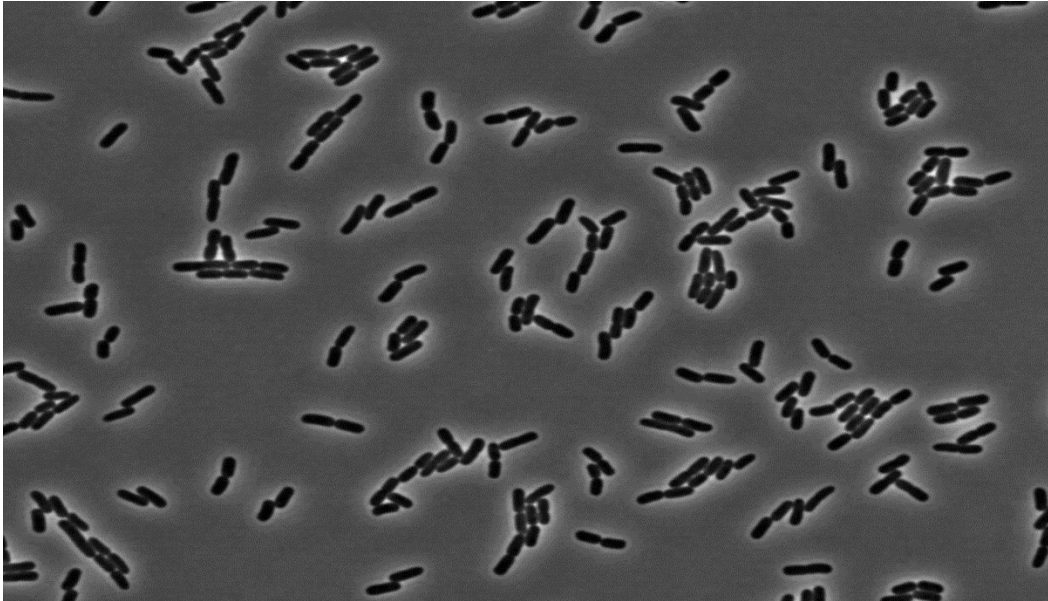


Figure 22: Single frame provided by MT. E. coli bacteria forming small micro-colonies

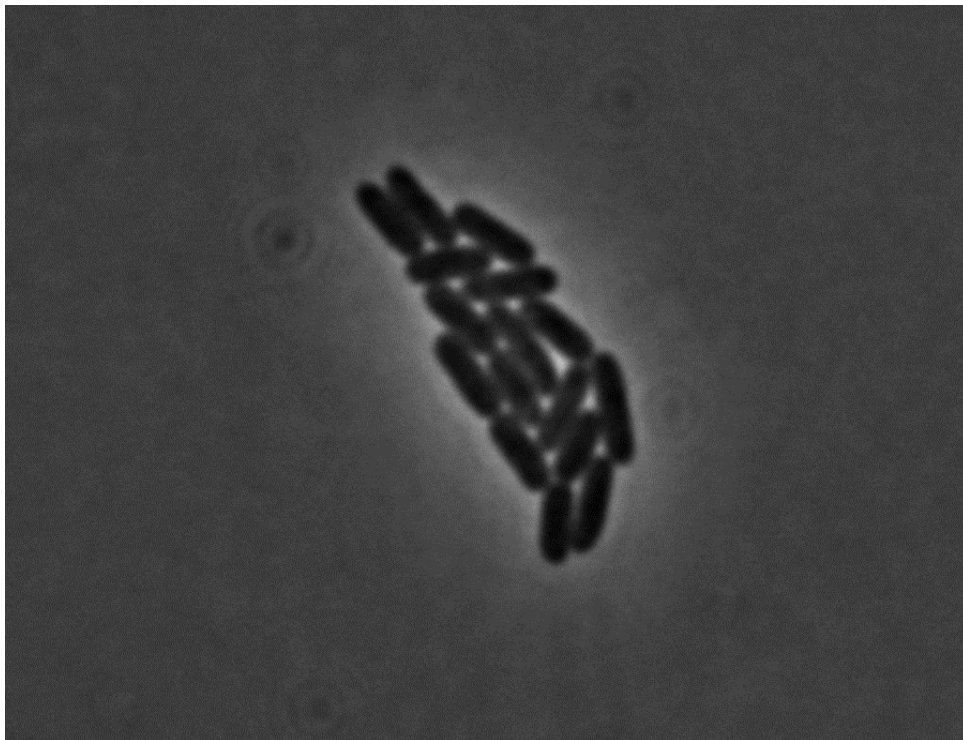


Figure 23: Single frame provided by CellTracer. E.coli bacteria forming a single micro-colony[13].

CellTracer's single frame holds a micro-colony of E.coli bacteria acquired by phase contrast light microscopy and was derived from a dataset produced by *Elowitz and Rosenfeld* [13] and it is available from CellTracer's webpage (please refer to Fig. 23 for details).

TLM-Tracker's image shows a micro-colony of *Bacillus megaterium* bacteria generated by *Stammen et al.* [12]. The image was derived from an image sequence which is available at TLM-Tracker's webpage. The dataset was generated by bright field optical microscopy (please refer to Fig. 24 for details).

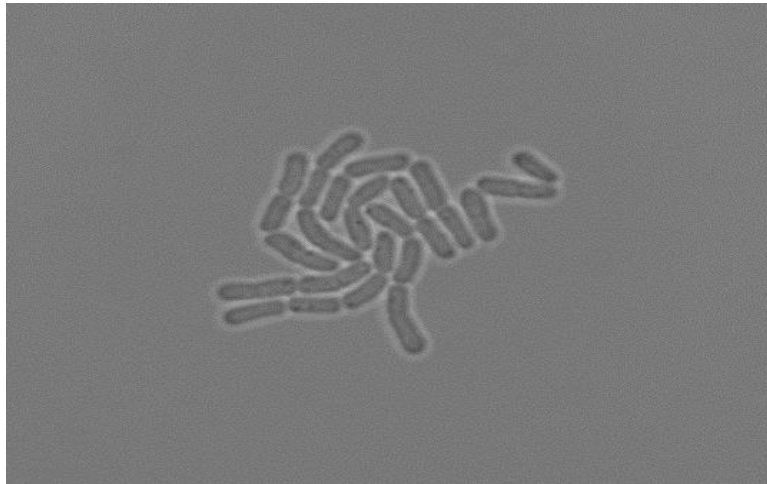


Figure 24: Single frame provided by TLM-Tracker [12]. *B. megaterium* bacteria forming a single micro-colony.

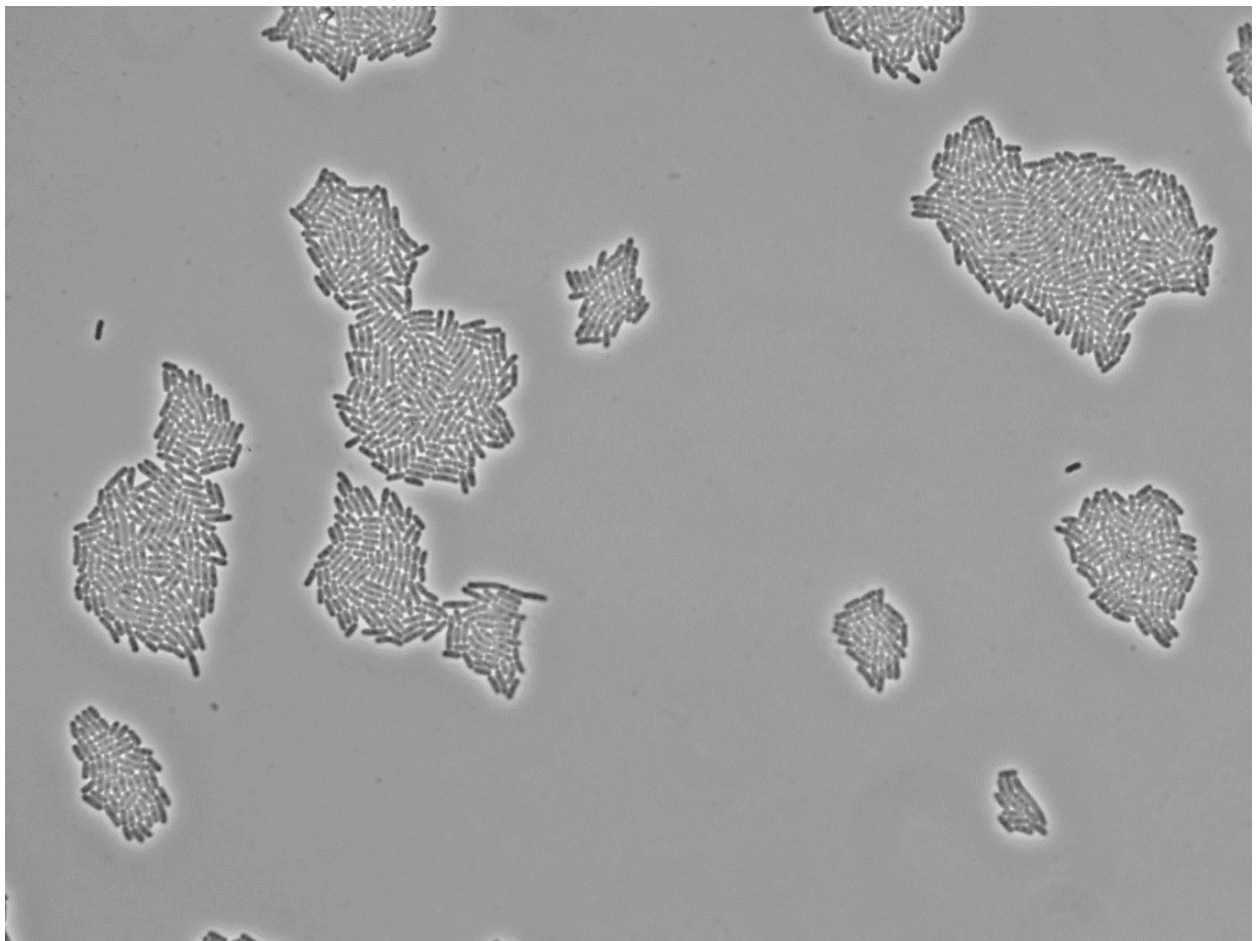


Figure 25: Single frame from multi-salPhase dataset. *S. typhimurium* bacteria forming multiple micro-colonies

Finally, in order to demonstrate the power of the proposed algorithm in processing large number of cells efficiently we used several frames of a dataset, named from now on as multi-salPhase, which contains multiple growing micro-colonies of salmonella (Fig. 25), acquired by phase contrast optical microscopy (please refer to [41] for further details).

The number of bacteria in some micro-colonies of this dataset exceeds the one thousand.

4.2 Evaluation Scheme

For the evaluation we used a two-way scheme as follows. First, we compare the counting results to the ground truth. The ground truth evaluation, as we shall call it from now on, was performed on the two time series datasets, salPhase and confocal GFP dataset. The ground truth of these datasets was created by manual counting and provided by expert users (please refer to Fig. 26 for details). Then we compare the cells counted by the proposed method to the results from the current state of the art software and to the ground truth. The comparative evaluation, as it will be called from now on, was performed on the single frame images described previously.

The evaluation was performed in terms of commonly used metrics, i.e. true positives (TPs) which are actual bacteria that were correctly classified as bacteria, false positives (FPs) which are artifacts that were incorrectly classified as bacteria and false Negatives (FNs) which are actual bacteria that were not classified as bacteria. We must note that artifacts can be either true image artifacts or segments of oversegmented bacteria. Furthermore we calculate several metrics that can exhibit the efficiency of the software. True positive rate (TPR) which represents the percentage of the actual bacteria in an image found by each method and defined as:

$$TPR = \frac{TP}{TP + FN}$$

Another metric is positive predictive value (PPV) which shows the probability that a detected bacterium is true positive. PPR is defined as it is illustrated below:

$$PPV = \frac{TP}{TP + FP}$$

The former is used to evaluate the sensitivity (recall) and the latter the precision of each method. We have also used the F-measure to assess the accuracy of each method. F-measure is a harmonic mean commonly used to characterize the sensitivity versus precision trade-off of competing methods, defined as:

$$F - measure = \frac{2 \cdot (TPF \cdot PPV)}{(TPF + PPV)}$$

In order to understand the evaluation scheme, we present an example of a single frame evaluation of salPhase dataset, in Fig. 27. Simultaneously, the segmentation results of each colony are illustrated. Going clockwise in each white panel and starting at the top left square, we represent the TP, FN and FP correspondingly,

Frame\Colony	1	2	3	4
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	1	1	1	1
⋮				
82	89	124	163	1
83	90	131	176	1

Figure 26: Example of a ground truth

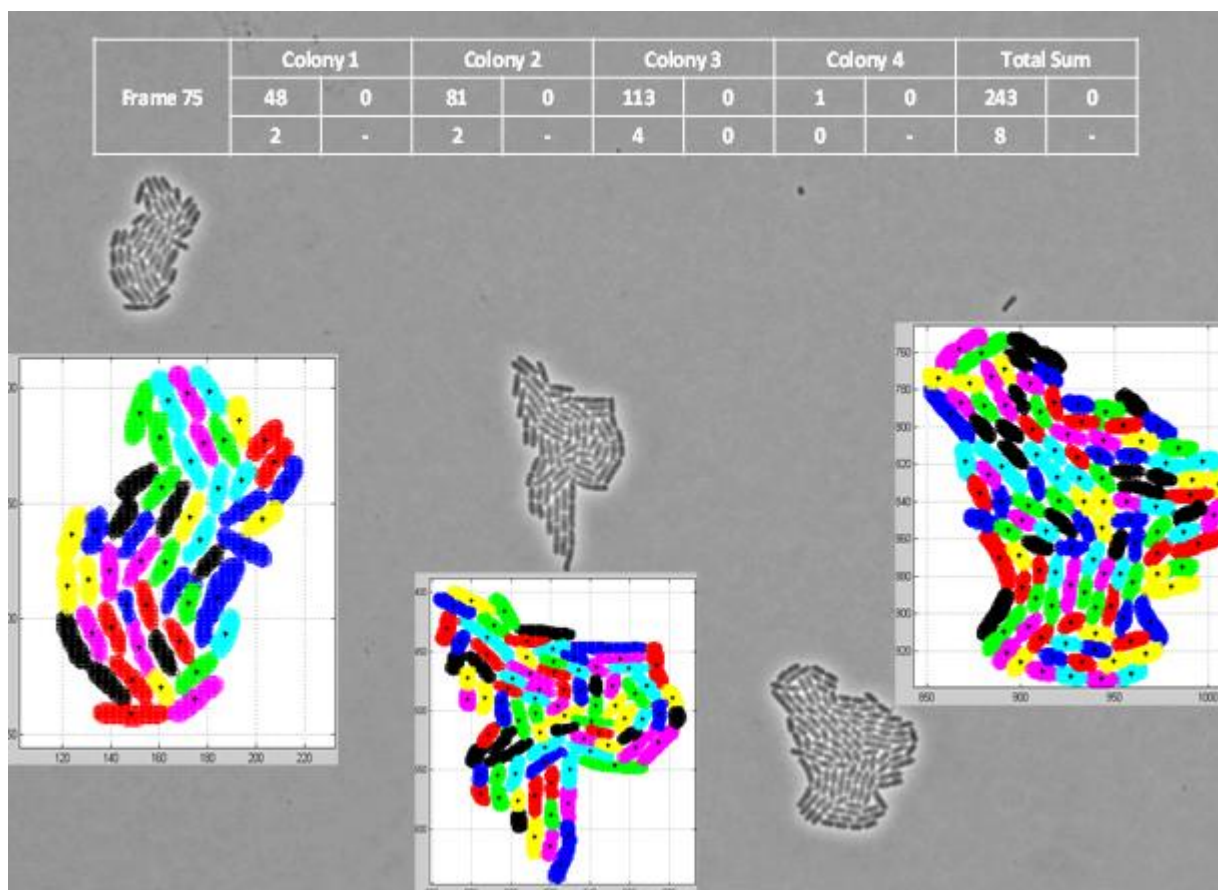


Figure 27: Demonstration of evaluation and segmentation results of a single frame of the salPhase dataset by the proposed methodology.

5. RESULTS & DISCUSSION

As we already stated in the Introduction Chapter, image processing and especially the segmentation of bacterial cells from microscopy images is the most critical task in microbial related community towards high throughput studies. It is still considered as a bottleneck in the experimental procedure of microbial studies since no really automation and consecutively high throughputness is achieved. In this chapter we will present the advantages of the developed methodology over current existing state of the art, following the aforementioned evaluation scheme.

5.1 Pipeline Evaluation based on Ground truth

In this section, we assess the robustness and efficiency of our pipeline with respect to different imaging modalities datasets. As mentioned before, we will use the salPhase and GFP confocal datasets of Koutsoumannis' lab and present the results in comparison to the ground truth.

Figure 28 presents a summary of the evaluation results versus ground truth. We can see that the developed method achieves a high recall (over 99%) and a high precision (over 96%) for either dataset. This can be translated as an indication that the developed methodology is both sensitive and precise. The importance of these results increases when someone considers the fact that the two datasets are generated by different types of microscopes (optical and confocal) and imaging method (phase contrast and differential interference contrast-DIC). In order to further support the results, we have also computed the F-measure which is used to characterize the sensitivity versus precision trade-off. The developed method achieves an F-measure of about 98% for both datasets, showing further its efficiency and robustness across imaging techniques. Additionally, in the same figure, we present the cumulative segmentation errors, i.e. the sum of FPs and FNs, for the two datasets. From the curves, one can observe that most of the errors appear after the mid time of each experiment. At that time the colonies are very crowded and the cells start to overlap and grow in 3-dimensions, leaving the 2-D structure of the colony. It is noticeable that we count 6856 bacteria out of 6895 counted manually (ground truth) for the salPhase dataset and only 263 cumulative errors. For the GFP confocal dataset, the pipeline counted 968 bacteria out of 969 counted by the annotators and only 40 cumulative errors.

Next, we move to visual inspection of the segmentation results. Figure 29 presents some segmentation examples of salPhase (line 1) and confocal GFP (line 2) datasets. As it can be observed from a visual review of the figure, the segmentation is pretty accurate and the overlaid red contours outline in detail the real contour of each bacterium (line 1 leftmost and line 2 leftmost). It is obvious that having such an accurate segmentation, we are able to have trustworthy feature extraction and the following downstream analysis (e.g. statistical analysis, model development and kinetics estimation). The two rightmost images of line 1 present a color-based visualization of the area, overlaid on the cells, and the corresponding histogram respectively. The different colors overlaid on the bacteria represent their size in squared micrometers, as indicated by the adjacent color bar. This visualization enables the user to easily distinguish bacteria independently from the feature chooses to visualize. The adjacent to the right image displays the cell area histogram. The two rightmost images at line 2 of Fig. 29 show the GFP quantification visualization and the corresponding histogram respectively. In the first image, the different colors of the bacteria represent their average GFP intensity included in each cell contour as color coded by corresponding color bar. The results are presented on a synthetic image based on the original input image.

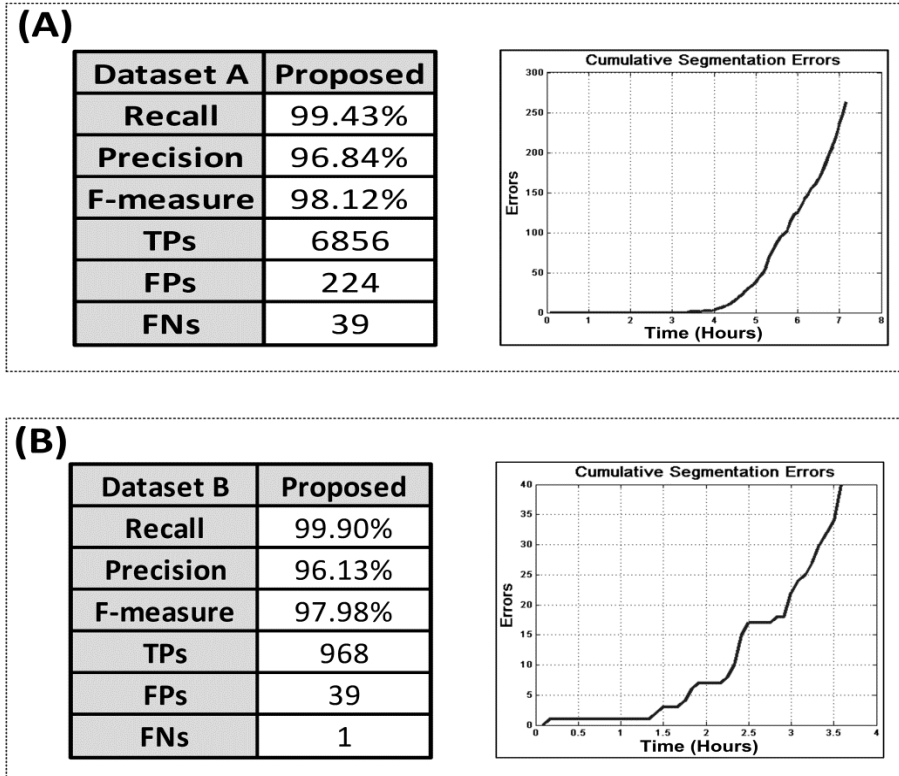


Figure 28: Evaluation results for SalPhase dataset. Panel A: Evaluation results for the complete dataset and cumulative segmentation errors (FNs+ FPs) (please refer to Chapter 4). Panel B: The same as in Panel A but for the confocal GFP dataset. As we can see, the developed method, while fully automated and user friendly, achieves a high F-measure score (above 97%). This measure is consistent regardless the image acquisition modality (i.e. optical, confocal, phase contrast /bright field/GFP staining).

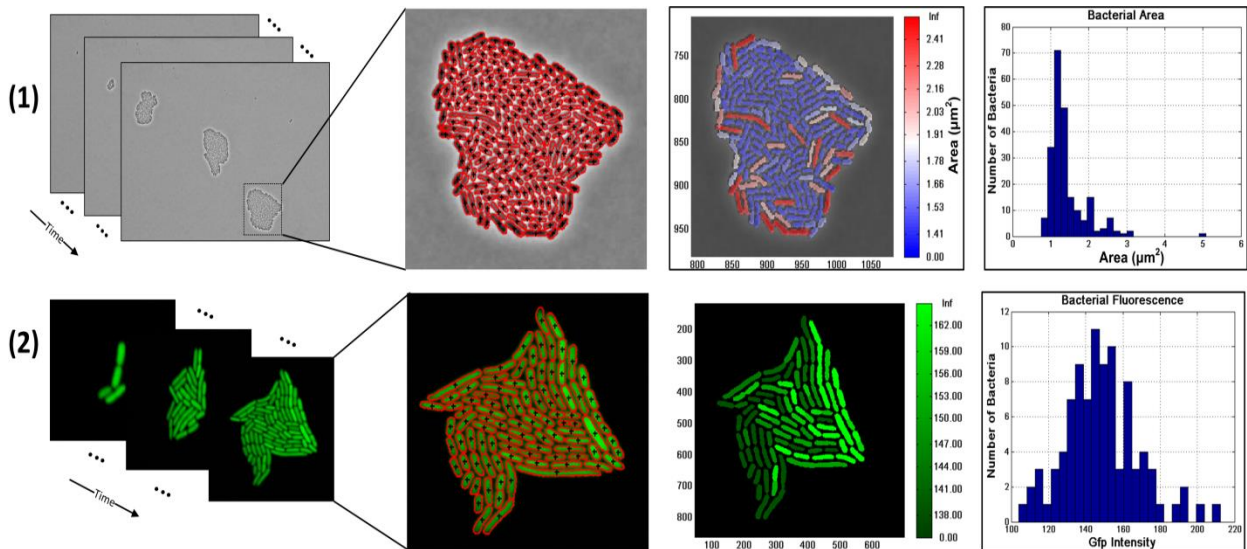


Figure 29: Evaluation results presentation for SalPhase dataset. Row 1 from left to right: Original dataset (stack of phase contrast images), cell segmentation result on one colony (red ellipses indicate the contours of segmented cells), and bacterial area visualization of the same colony and histogram of the segmented bacterial areas. Row 2 from left to right: Original dataset (stack of confocal GFP images), cell segmentation result (red ellipses indicate the contours of segmented cells), GFP visualization of the colony on the initial image and histogram of the bacterial GFP quantity. We can observe from the figure that the developed methodology is efficient and robust while fully automated for a stack of images, usually exhibiting limited intra-variability in terms of focus and illumination.

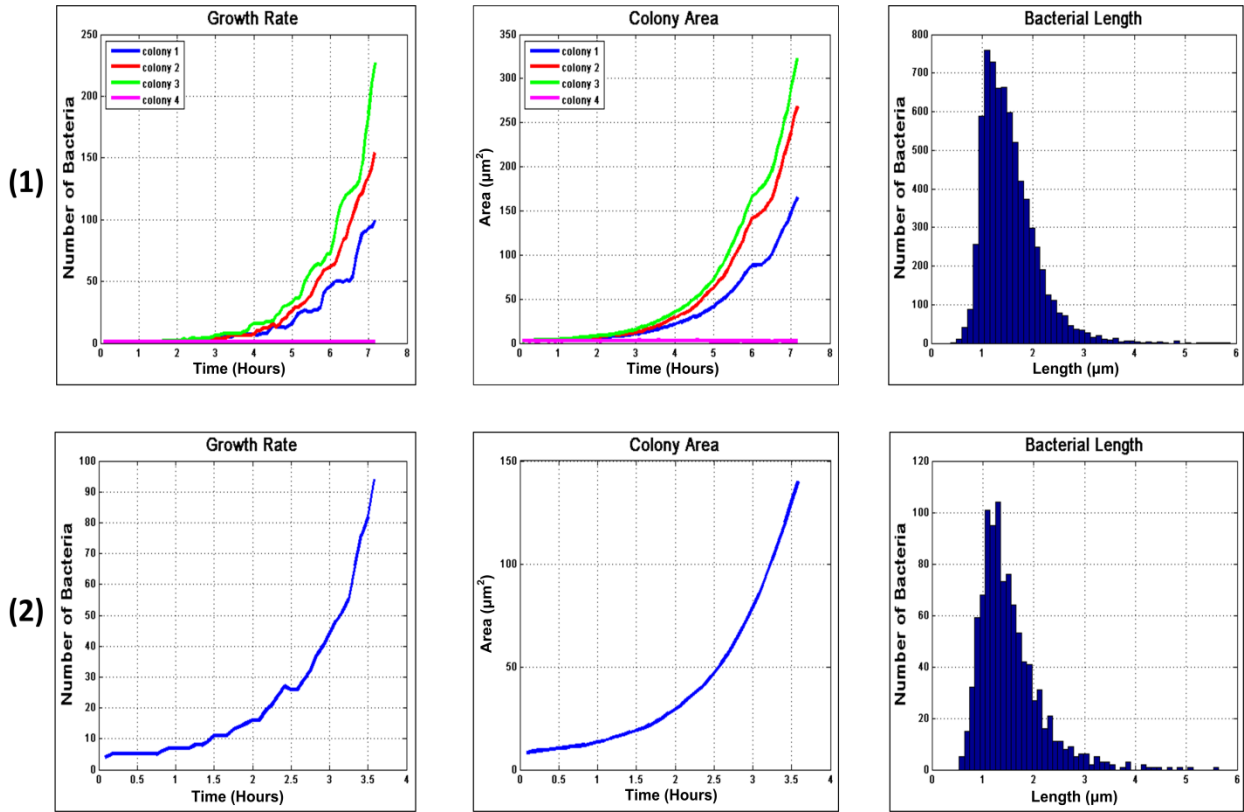
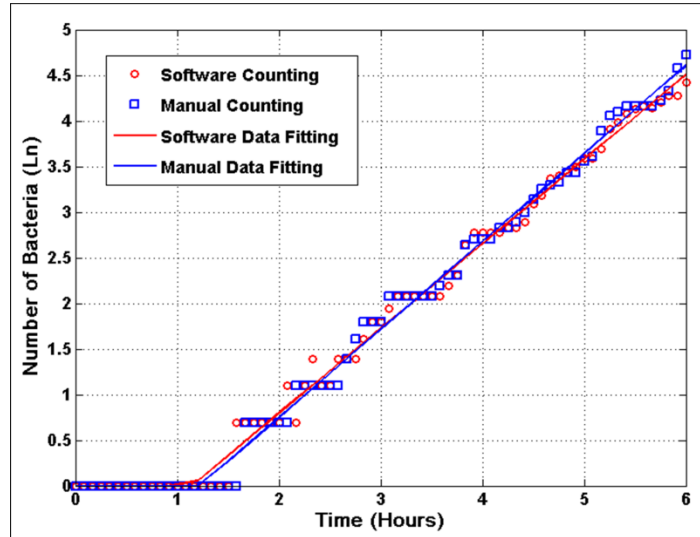


Figure 30: Bacterial Growth curves for each colony in the salPhase dataset (line 1) and in the GFP confocal dataset (line 2), corresponding colony Area Growth curve, and bacterial length distribution. We can observe from the figure that the developed methodology embeds the capability of measuring several colony and single cell properties of high importance for further analysis.

Figure 30 presents several software capabilities offered for further data analysis of the results. Specifically, the subfigures refer to the salPhase and confocal dataset correspondingly. The leftmost graph presents the growth rate of each colony individually for the 86 frames of the dataset (approximately 7.17 hours). It can be clearly observed that the number of bacteria in each colony increases exponentially as it is expected according to [41]. The same holds for the colony area. Additionally, the developed software can produce histograms for all the single cell features that were mentioned in Section 3.5. For example, we present here the histogram of bacterial length (in micrometers) of a growing colony which can be used to infer the corresponding distribution. Another interesting feature is that one can also create movies of growing colonies visualizing single cell features.

As an additive validation and presentation of the developed method in terms of systems biology point of view, we fitted the bacterial number increase with time for each micro-colony to the primary model of *Baranyi and Roberts* [43] for the estimation of the growth kinetic parameters lag time (λ) and specific maximum growth rate (μ_{max}). In order to describe the abrupt transition from the lag to the exponential phase characterizing the observed growth, the values of the parameters m and n of the model were fixed to 0 and 20, respectively as done in [41]. Figure 31 shows a comparison of the fittings for data counted manually and with the developed software for a micro-colony and the estimated kinetic parameters for the three micro-colonies. As one can observe, for all micro-colonies tested, the kinetic parameters that were estimated from the manually counted bacteria are almost identical to those estimated from data derived from the



Manual Counting				
Colony	Rate (h-1)	Lag (h)	Se (fit)	R ² _stat
1	0.808	1.373	0.153	0.991
2	1.038	1.846	0.108	0.995
3	0.965	1.225	0.120	0.994

Software Counting				
Colony	Rate (h-1)	Lag (h)	Se (fit)	R ² _stat
1	0.8134	1.377	0.1416	0.9916
2	1.031	1.865	0.1115	0.9949
3	0.9279	1.143	0.1206	0.9935

Figure 31: Comparison of the fittings for salPhase dataset counted manually and by the developed software for a micro-colony; Tables present the kinetic parameters of microbial growth estimated.

developed software. These results indicate the software can provide accurate data for predictive microbiology as it is outlined in [41].

Conclusively, the developed methodology is efficient and robust while fully automated even for stacks of images, usually exhibiting limited intra-variability in terms of focus and illumination. This intra-variability of a dataset, although limited, has tremendous impact on image analysis since standard techniques are often so sensitive that output no results at all. Furthermore, it embeds the capability of measuring several colony and single cell properties of high importance for further analysis.

5.2 Software Evaluation versus State-of-the-Art

At this point, we would like to justify the choice of the specific state of the art software packages used in the conducted comparative evaluation. Those software packages are the most commonly used in microbiological labs worldwide. As such, they have been used in several microbial studies [44-50]. Our objective here is to show that the

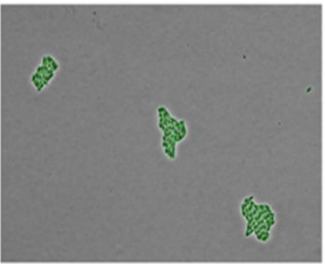
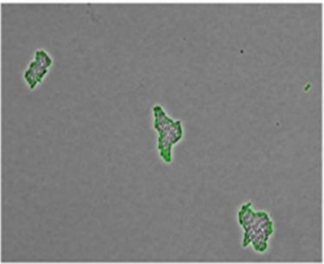
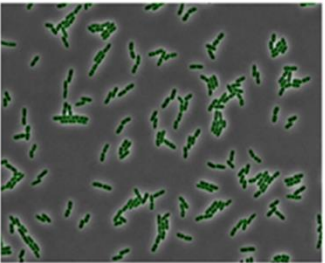
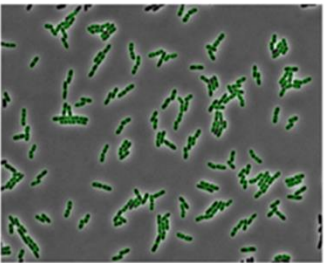
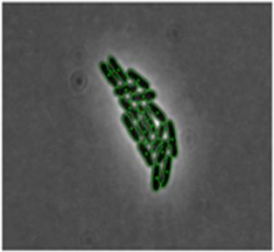
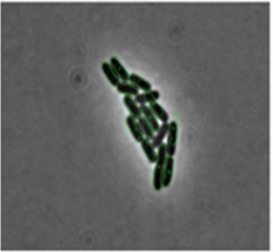
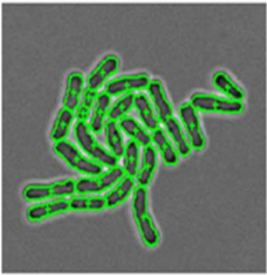
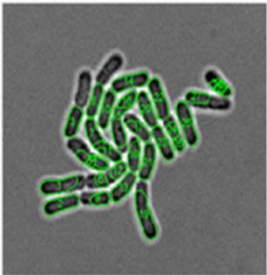
	Proposed	MicrobeTracker	Evaluation Summary																																			
Proposed Dataset (1)			<table border="1"> <thead> <tr> <th>Method</th> <th>TPs</th> <th>FPs</th> <th>FNs</th> <th>Recall</th> <th>Precision</th> <th>F-measure</th> </tr> </thead> <tbody> <tr> <td>Proposed</td> <td>128</td> <td>6</td> <td>1</td> <td>99.22%</td> <td>95.52%</td> <td>97.34%</td> </tr> <tr> <td>MicrobeTracker</td> <td>41</td> <td>0</td> <td>88</td> <td>31.78%</td> <td>100.00%</td> <td>48.24%</td> </tr> <tr> <td>CellTracer</td> <td colspan="6" style="text-align: center;">-</td> </tr> <tr> <td>TLM Tracker</td> <td colspan="6" style="text-align: center;">-</td> </tr> </tbody> </table>	Method	TPs	FPs	FNs	Recall	Precision	F-measure	Proposed	128	6	1	99.22%	95.52%	97.34%	MicrobeTracker	41	0	88	31.78%	100.00%	48.24%	CellTracer	-						TLM Tracker	-					
Method	TPs	FPs	FNs	Recall	Precision	F-measure																																
Proposed	128	6	1	99.22%	95.52%	97.34%																																
MicrobeTracker	41	0	88	31.78%	100.00%	48.24%																																
CellTracer	-																																					
TLM Tracker	-																																					
MicrobeTracker's Dataset (2)			<table border="1"> <thead> <tr> <th>Method</th> <th>TPs</th> <th>FPs</th> <th>FNs</th> <th>Recall</th> <th>Precision</th> <th>F-measure</th> </tr> </thead> <tbody> <tr> <td>Proposed</td> <td>246</td> <td>0</td> <td>3</td> <td>98.80%</td> <td>100.00%</td> <td>99.39%</td> </tr> <tr> <td>MicrobeTracker</td> <td>243</td> <td>0</td> <td>6</td> <td>97.59%</td> <td>100.00%</td> <td>98.78%</td> </tr> <tr> <td>CellTracer</td> <td colspan="6" style="text-align: center;">-</td> </tr> <tr> <td>TLM Tracker</td> <td colspan="6" style="text-align: center;">-</td> </tr> </tbody> </table>	Method	TPs	FPs	FNs	Recall	Precision	F-measure	Proposed	246	0	3	98.80%	100.00%	99.39%	MicrobeTracker	243	0	6	97.59%	100.00%	98.78%	CellTracer	-						TLM Tracker	-					
Method	TPs	FPs	FNs	Recall	Precision	F-measure																																
Proposed	246	0	3	98.80%	100.00%	99.39%																																
MicrobeTracker	243	0	6	97.59%	100.00%	98.78%																																
CellTracer	-																																					
TLM Tracker	-																																					
CellTracer's Dataset (3)			<table border="1"> <thead> <tr> <th>Method</th> <th>TPs</th> <th>FPs</th> <th>FNs</th> <th>Recall</th> <th>Precision</th> <th>F-measure</th> </tr> </thead> <tbody> <tr> <td>Proposed</td> <td>16</td> <td>1</td> <td>0</td> <td>100.00%</td> <td>94.12%</td> <td>96.97%</td> </tr> <tr> <td>MicrobeTracker</td> <td>11</td> <td>0</td> <td>5</td> <td>68.75%</td> <td>100.00%</td> <td>81.48%</td> </tr> <tr> <td>CellTracer</td> <td>15</td> <td>0</td> <td>1</td> <td>93.75%</td> <td>100.00%</td> <td>96.77%</td> </tr> <tr> <td>TLM Tracker</td> <td>13</td> <td>2</td> <td>3</td> <td>81.25%</td> <td>86.67%</td> <td>83.87%</td> </tr> </tbody> </table>	Method	TPs	FPs	FNs	Recall	Precision	F-measure	Proposed	16	1	0	100.00%	94.12%	96.97%	MicrobeTracker	11	0	5	68.75%	100.00%	81.48%	CellTracer	15	0	1	93.75%	100.00%	96.77%	TLM Tracker	13	2	3	81.25%	86.67%	83.87%
Method	TPs	FPs	FNs	Recall	Precision	F-measure																																
Proposed	16	1	0	100.00%	94.12%	96.97%																																
MicrobeTracker	11	0	5	68.75%	100.00%	81.48%																																
CellTracer	15	0	1	93.75%	100.00%	96.77%																																
TLM Tracker	13	2	3	81.25%	86.67%	83.87%																																
TLM Tracker's Dataset (4)			<table border="1"> <thead> <tr> <th>Method</th> <th>TPs</th> <th>FPs</th> <th>FNs</th> <th>Recall</th> <th>Precision</th> <th>F-measure</th> </tr> </thead> <tbody> <tr> <td>Proposed</td> <td>25</td> <td>1</td> <td>0</td> <td>100.00%</td> <td>96.15%</td> <td>98.04%</td> </tr> <tr> <td>MicrobeTracker</td> <td>22</td> <td>7</td> <td>3</td> <td>88.00%</td> <td>75.86%</td> <td>81.48%</td> </tr> <tr> <td>CellTracer</td> <td colspan="6" style="text-align: center;">-</td> </tr> <tr> <td>TLM Tracker</td> <td>23</td> <td>0</td> <td>2</td> <td>92.00%</td> <td>100.00%</td> <td>95.83%</td> </tr> </tbody> </table>	Method	TPs	FPs	FNs	Recall	Precision	F-measure	Proposed	25	1	0	100.00%	96.15%	98.04%	MicrobeTracker	22	7	3	88.00%	75.86%	81.48%	CellTracer	-						TLM Tracker	23	0	2	92.00%	100.00%	95.83%
Method	TPs	FPs	FNs	Recall	Precision	F-measure																																
Proposed	25	1	0	100.00%	96.15%	98.04%																																
MicrobeTracker	22	7	3	88.00%	75.86%	81.48%																																
CellTracer	-																																					
TLM Tracker	23	0	2	92.00%	100.00%	95.83%																																

Figure 32: Software comparison evaluation with the state-of-the-art. First two columns present the segmentation/cell detection results of the developed and MicrobeTracker. For the CellTracer and TLM Tracker we show no segmentation images since they could not operate with all different datasets. The third column summarizes the performance of each method for every different dataset. As we can see from the segmentation results (first column) and the relative F-measure ($\geq 96\%$ for all cases), the proposed method is robust for different imaging modalities (optical and confocal phase contrast, optical bright field) and for data acquired by different labs. Furthermore, if one inspects the segmented images, it is clear that the cells' contours are more reliable, leading to a more efficient and robust cell property estimation and GFP measurements.

proposed pipeline is an improvement over the state of the art and to exhibit its ability to process images from different laboratories and imaging methods.

In order to do so, we used the same data that were used by the state-of-the-art software to demonstrate their performance upon their presentation. For the sake of an objective evaluation, we evaluated each software package with all considered datasets. At this

point we must stress that the state of the art software packages need different parameterization for images acquired by different imaging methods. Thus, we were forced to do extensive experimentation so as to choose the “optimal” parameters set. We must also state that the aforementioned parameterization is not trivial and straightforward since it also requires to some extent, knowledge of image processing concepts (i.e. filter types and their corresponding properties, image segmentation methods, etc). On the other hand, the parameterization of the proposed method is independent of image modality and depends only to image resolution. The parameters are set automatically by the software, once the user inputs the resolution, something that frees the user from image processing related concepts.

Figure 32 exhibits the comparative evaluation results. The *microbeTracker* was found to be the most robust since it is the only that was able to produce results for all images used in the evaluation. The rest of the packages gave no results for some image modalities despite that we tried several different parameterization. The first and the second column of Fig. 32 present the segmentation results (the detected cells are outlined by green ellipse) of the developed method and *microbeTracker*.

It is clear that the developed methodology is superior from *MicrobeTracker* in all cases. We can notice that in Fig. 32 line 1 column 2, *MicrobeTracker* gives extremely inaccurate results. One could say that this is prospectively since this dataset is ours but this claim is rejected by the rest outcomes in the evaluation. So, in order to quantify the performance of the different methods we present at the third column of Fig. 32, the corresponding summary tables. The proposed method exhibits a notable F-measure advantage (over 96%) for all datasets considered. The most significant finding is that the proposed pipeline outperforms the rest of the software packages even when it comes to their own images. The dashes shown in the tables indicate that the specified software did not give results for the specific image modality. The software that was found to be the most specific in terms of imaging modality is *CellTracer* which gave results only for its own dataset. It is remarkable that no software package except for the proposed gave accurate result in *salPhase* dataset which contains more than one and even over-crowded micro-colony. Additionally *CellTracer* and *TLM Tracker* cannot analyze images with sparse bacteria and small micro-colonies, such as *MicrobeTracker*'s image. In summary, we can see that the proposed pipeline found approximately all the bacteria (for all the images under consideration), i.e. high recall percentage-over 98%, while traced a few artifacts as bacteria, i.e. high precision percentage-over 94%. At this point and after this evaluation we are safe to conclude that the developed methodology not only exhibits superior performance to the current state of the art but also is the most generic one since it can process and give trustworthy results for several combinations of laboratories, microscopy types and imaging techniques. Finally, the developed methodology is fully automated, since does not require the user to be familiar with advanced image processing concepts nor requiring obscure input rather than the dataset to process and the relationship of physical dimensions to imaging pixel size (this information is given by the image acquiring software of the microscope system).

5.3 Discussion and Additional Software Capabilities

Cell lineage construction is a profoundly significant functionality that can favor microbiologists in their research. Its importance lies on the fact that the information contained can be used to track specific features of a single cell from the time of its birth to several progenitors after and several ancestors before. *CellTracer* and *TLM-Tracker* provide this capability although they do not exploit extensively the derived information,

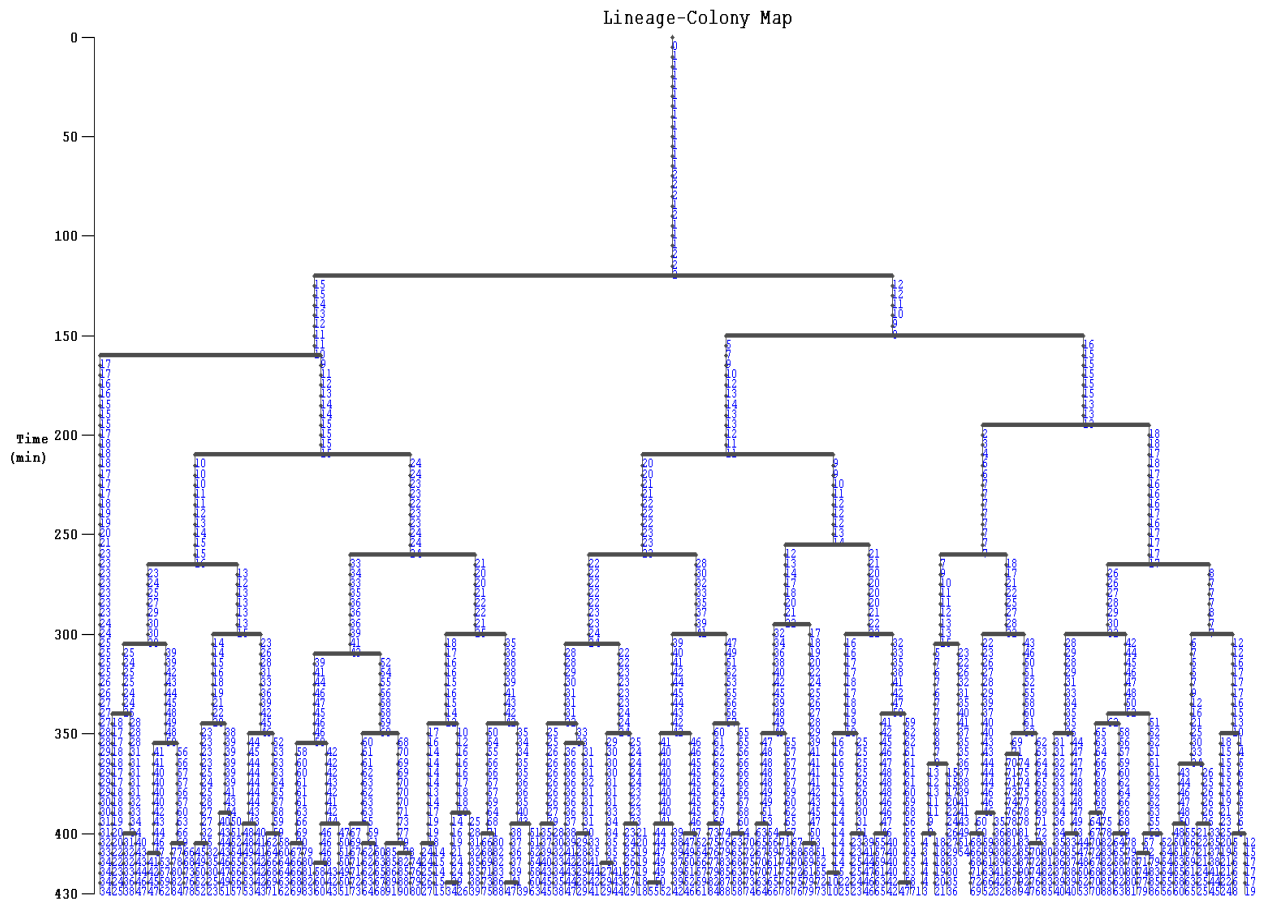


Figure 33: Relative cell position to the centroid of the colony (colony mapping).

while MicrobeTracker does not. In the developed software package, the user is offered the ability to link to the cell lineage with several extracted features during the dataset processing, while all the aforementioned visualization capabilities are governed by user-defined queries. The displayed trees in Fig. 33 and Fig. 34 correspond to the same cell lineage through time (left vertical axis) from salPhase dataset while the displayed trees of Fig. 35 and Fig. 36 correspond to the same cell lineage through time (left vertical axis) from GFP confocal dataset. Each figure shows one different visualization capability available from the software. One interesting feature that can be visualized is the position of each bacterium relatively to the colony's centroids (Fig. 33).

Figure 34 shows the cells' area lineage of salPhase dataset's first colony. One can easily see how the area of a cell is growing until a division takes place. When a node turns from red to blue and then again to red, it is indicated that a segmentation inaccuracy occurs. Specifically, in the area visualization lineage, it is implied that the algorithm made a segmentation error and then estimated the bacterial area wrongfully. In Figure 35 is illustrated the GFP quantification on the lineage of confocal GFP dataset's colony. We must note that this functionality is offered by TLM-Tracker too. Nonetheless, none of the competitor software packages infer the division time of the bacteria.

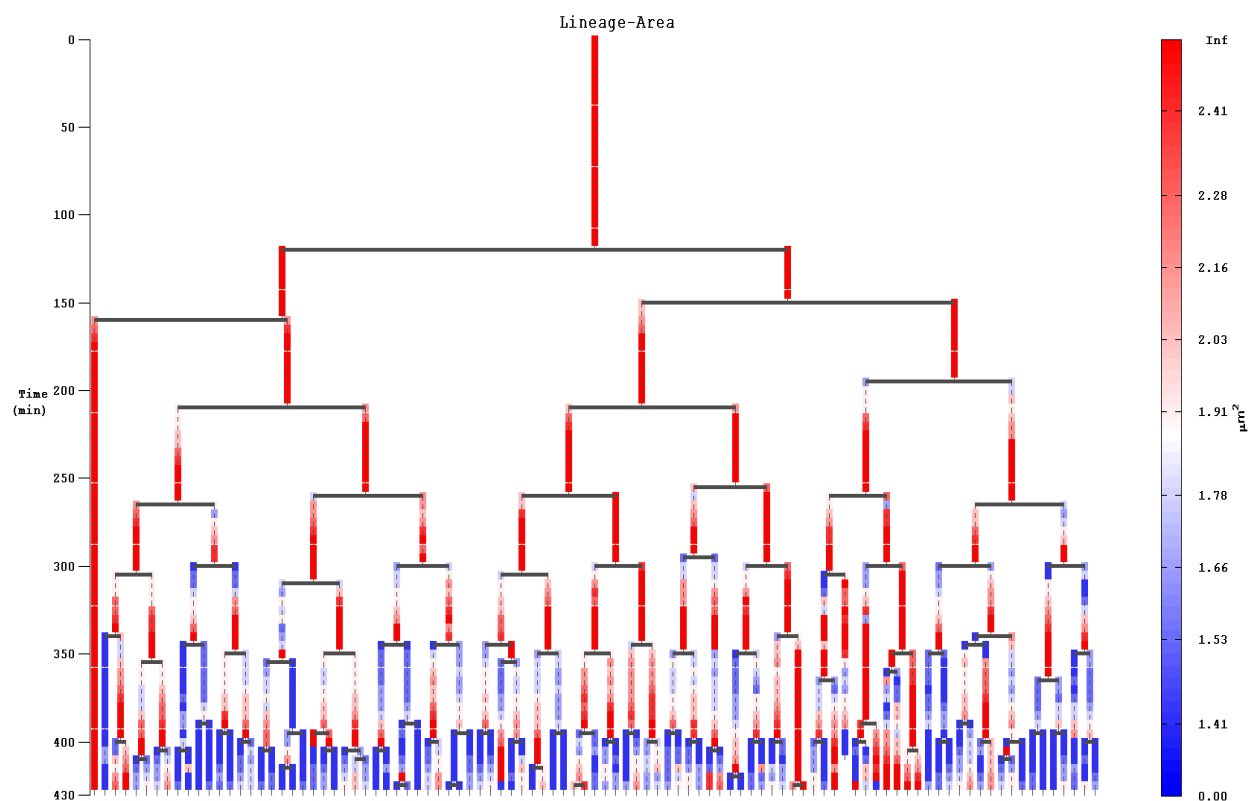


Figure 34: Area visualization, we can see the evolution of cell area through time and cell lineage simultaneously, one easily see the “critical” cell size just before cell division.

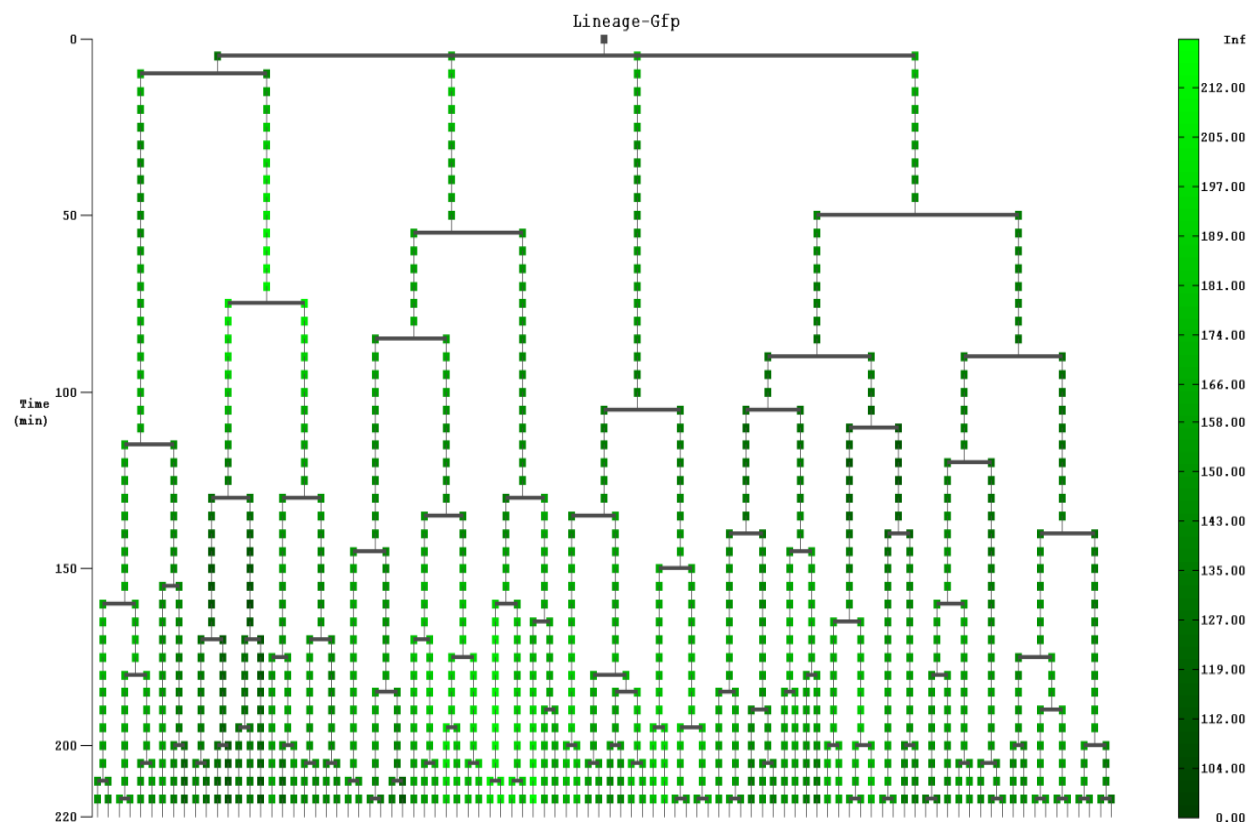


Figure 35: GFP quantification visualization through time (GFP confocal dataset). The root node of the lineage is a pseudo node, i.e. they do not represent an actual bacterium, because the first frame of the dataset contains a micro-colony of four bacteria and not a single cell.

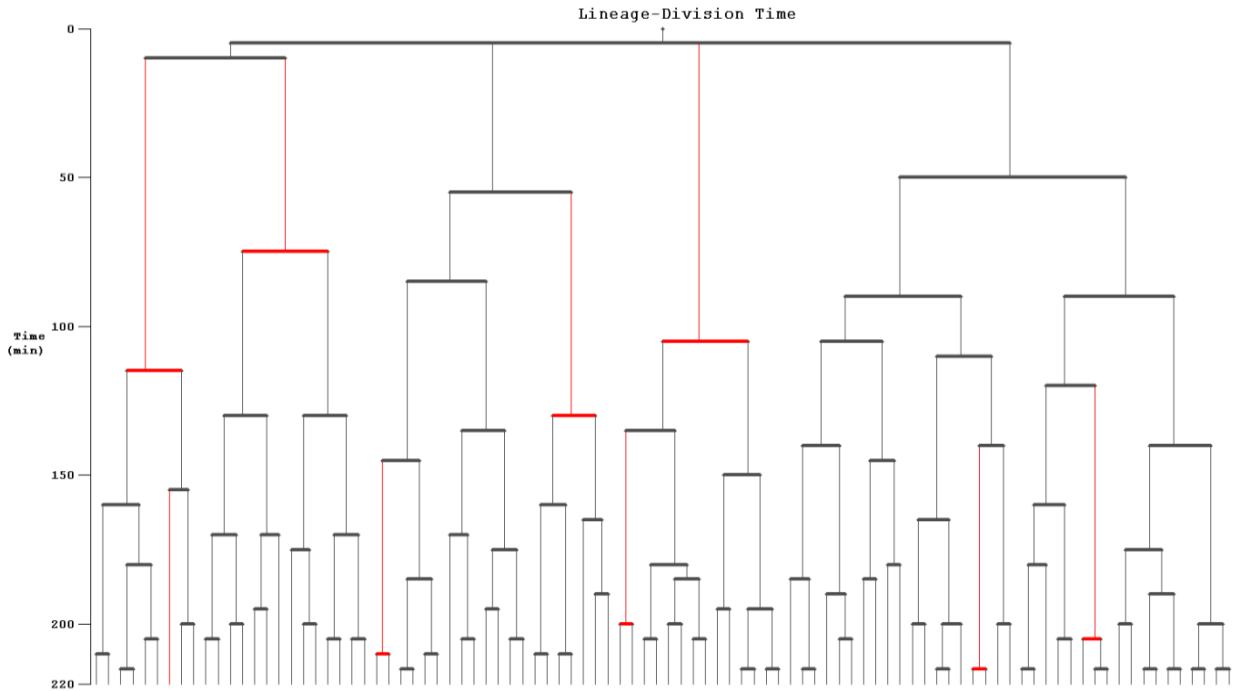


Figure 36: Cell division time visualization (red branch shows that the division time of a cell is more than 60 minutes). The root node of the lineage is a pseudo node, i.e. they do not represent an actual bacterium, because the first frame of the dataset contains a micro-colony of four bacteria and not a single cell.

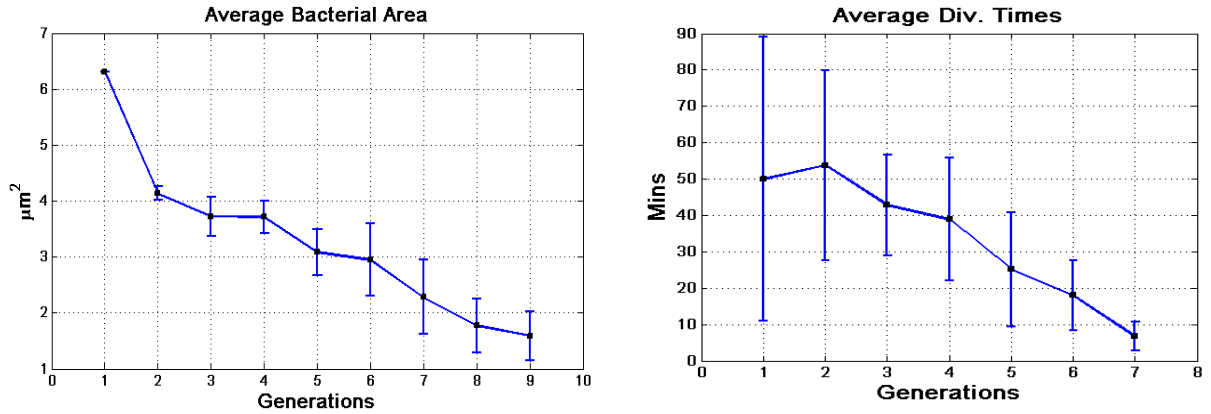


Figure 37: Average bacterium area per generation according to lineage of Fig. 33 (left). Average bacterium division time per generation according to lineage of Fig. 35 (right).

In Figure 36 we display the lineage which is linked to the division times of the cells. This is a very useful property since it can be used to record and study the division time across experiments under different conditions. An interesting application could be the linkage of division time with position of the cells in the growing colony. In addition, the proposed pipeline can mark the bacteria of the given lineage having division times between user defined thresholds (e.g. at Fig. 36 red branches show if the division time of a cell is more than 60 minutes).

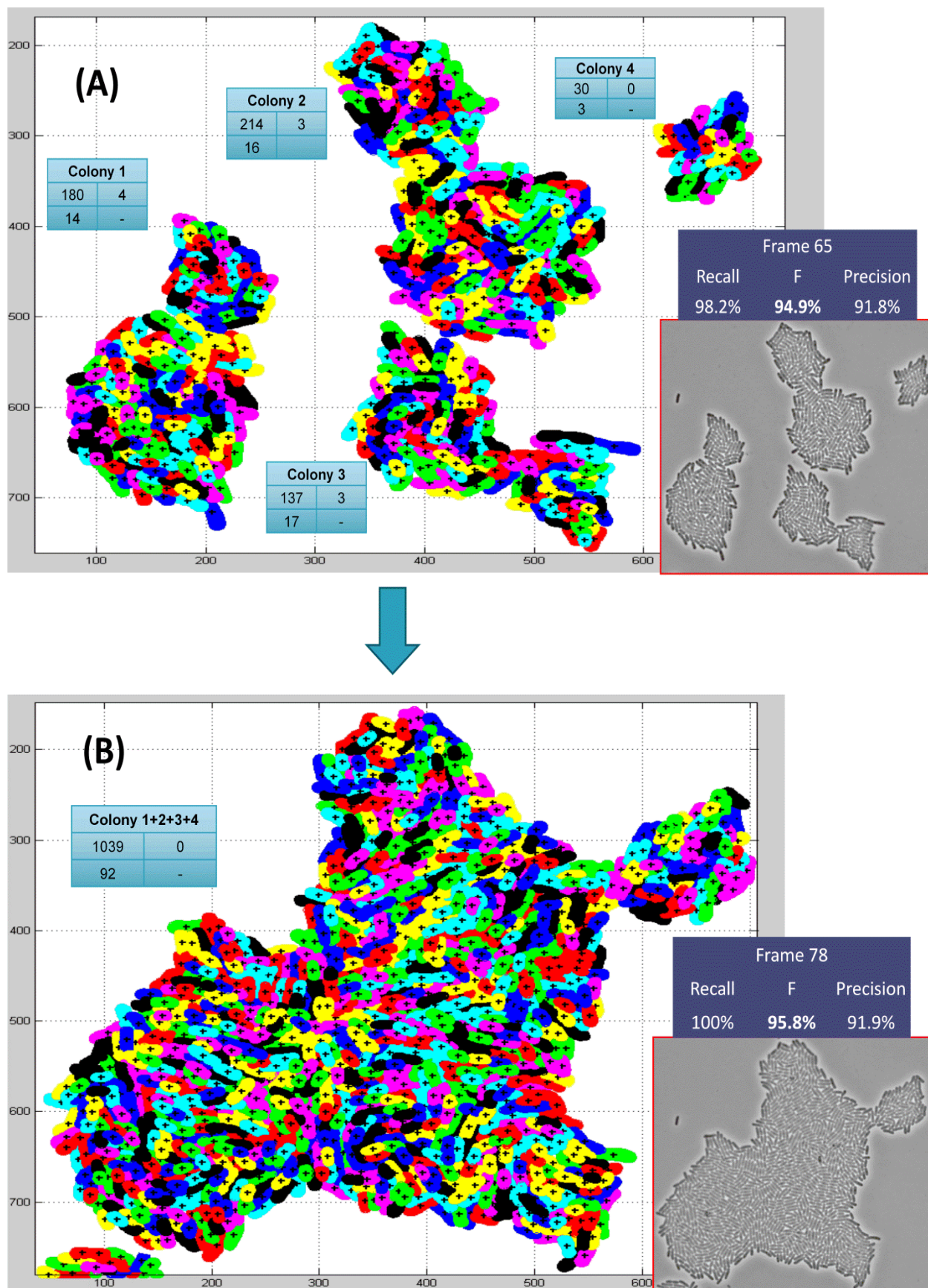


Figure 38: Multi-salPhase dataset analysis (frame 65 and 78). The four colonies merged with each other. The rightmost red frames illustrate the input images. The cyan and the purple boxes illustrate the evaluation statistics of each colony and each frame correspondingly.

Another novel functionality of the proposed methodology is that having the cell lineage we can track the cell generations in time. Knowing a single cell's generations we can create graphs of the average value (and relative deviations) of any (of the measured) feature per generation and histograms for each generation individually. For example, in Fig.37 left we show the average bacterial area for each generation of salPhase dataset, while in Fig. 37 right the average division time is displayed. To the best of our knowledge, the developed methodology is the only that offers this variety of capabilities as far as the lineage construction is concerned.

Another significant issue we should discuss is that the proposed methodology is able to analyze over-crowded images. As it was presented in Section 5.1, the algorithm can segment images with more than one colony in contrast to the rest of the software packages. However, the proposed pipeline is efficient even for analyzing over-crowded colonies, i.e. containing over thousand bacteria. It is also remarkable that for some datasets in which the micro-colonies merge with each other, the algorithm continues to segment efficiently the bacterial cells efficiently. Fig. 38 illustrates this functionality. We also provide the evaluation results of each colony individually for each frame. It can be observed that evaluation metrics of these images and the evaluation metrics of less difficult datasets (Section 5.1) are almost stable (recall over 98%, precision over 91% and F-measure over 95%).

Furthermore, we examined the proposed pipeline in terms of computational time. In general, the execution time of the proposed pipeline increases quadratically with the number of bacteria as implied by the left graph of Fig. 39. In this graph, execution time is illustrated versus time (where time is represented by frames) (salPhase dataset's images). As we reach the end of the dataset, the images contain more bacteria. In order to make a more precise analysis, we examined the segmentation time according to the number of bacteria. Practically, we subtracted the preprocessing time from the execution time, so we isolated the so called segmentation time. We can observe from the Fig. 39right graph that segmentation time is $O(n^2)$ where n stands for number of bacteria. The curve fluctuates because the segmentation time is affected by image's quality. If a specific frame is of less quality (e.g. due to noise existence), then oversegmentation problem increases, i.e. the initial segmentation is not so precise, so the watershed transformation segments to more sub-objects a colony's complex objects, and more time is needed for puzzle solving process. Consequently, we conclude that the algorithm complexity is quadratic not to the number of bacteria but to the number of fragments lying in each object. If we assume that image quality in a given dataset is constant we can infer computational time to increase analogous to the length of the colony. Simultaneously, the computational time of EM algorithm is pseudo-quadratic, having the iterations number of the algorithm to be constant, to the number of clusters (i.e. objects-bacteria) and the generated dataset's size, i.e. $O(k \cdot n)$. Thus, the time consumed by EM phase depends of a colony's size, i.e. while colony's size increases, k and n increases too. Consequently, the algorithm's complexity is quadratic. It is remarkable though that despite the fact computational time rises quadratically the computational time per bacterium is approximately constant, i.e. $O(1)$, see Fig. 39 middle graph for details.

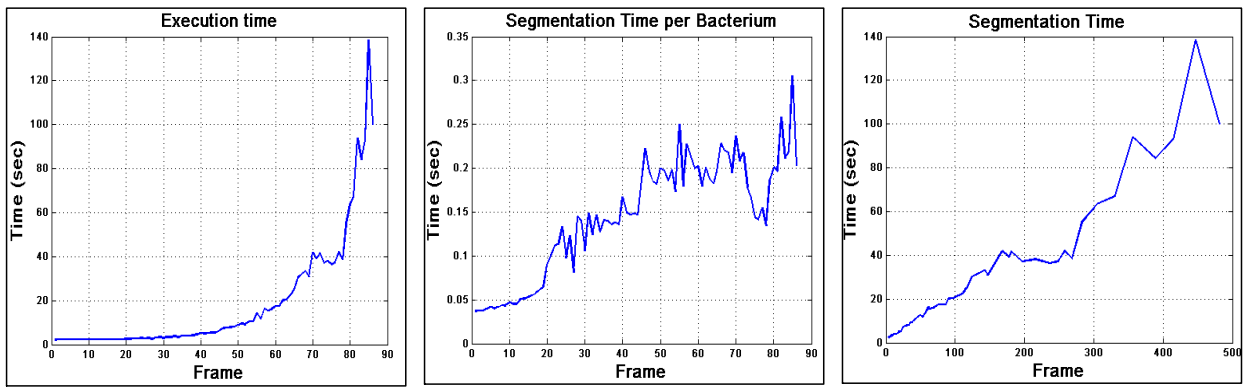


Figure 39: Computational time curves. Execution time of each frame (left). Segmentation time per bacterium of each frame (middle). Segmentation time of each frame (right).

6. CONCLUSION AND FUTURE WORK

In this work, we presented a fully automated tool, which can be found very useful while enabling high throughput processing of time lapse datasets in microbiology. The examples of the software utilization proved the robustness and universality of the proposed pipeline. The proposed methodology enables studying how biofilms are established and evolving under different environmental conditions in a system biology approach, by providing large amount of data about the development of bacterial colonies, i.e.:

- How fast a colony grows?
- How fast a colony's bacteria are duplicated?
- Check whether the position of a bacterium in the colony plays a significant role to its reproduction
- How environmental conditions affect a specific bacterial type and to what extent?

All the extracted information can be exploited so as to parameterize simulation models effectively. So our efficient high throughput analysis tool extracting several features from individual cells or complete micro-colonies, e.g. shape, area and axes length, is absolutely necessary.

Future work includes the evolution of the methodology focused on certain stages of the pipeline. Specifically, one improvement might be the skeleton extraction algorithm in order to derive more precisely results. So, by improving this, we will be able to compute the "deep valley" criterion more accurately. Additionally, improvements on the puzzle solving algorithm would make the cell synthesis from fragments more efficient. As far as, the lineage construction is concerned, we are going to advance the algorithm in order to become more insensitive to bacterial perturbations (inside a colony) and movement. This will make lineage construction algorithm more robust and reliable. Another promising extension, we want to accomplish is to analyze datasets of different environments, i.e. micro-fluidic devices or synthetic tissues. However, the most significant goal we want to achieve is to exploit the information and the data extracted from several datasets, i.e. datasets with different environmental conditions, in order to apply a system's biology approach in micro-colonies development.

REFERENCES

- [1] S. V. Avery, "Microbial cell individuality and the underlying sources of heterogeneity," *Nat. Rev. Microbiol.*, vol. 4, no. 8, pp. 577-587, Aug. 2006.
- [2] P. Gilbert, J. R. Das, M. V. Jones, and D. G. Allison., "Assessment of resistance towards biocides following the attachment of micro-organisms to, and growth on, surfaces," *J. Appl. Microbiol.*, vol. 91, no. 2, pp. 248-255, Aug. 2001.
- [3] P. Gilbert, D. G. Allison, and A. J. McBain, "Biofilms in vitro and in vivo: do singular mechanisms imply cross-resistance?," *J. Appl. Microbiol.*, vol. 92 Supp, pp. S98-S110.
- [4] B. B. Christensen, J. A. Haagensen, A. Heydorn, and S. Molin, "Metabolic commensalism and competition in a two-species microbial consortium," *Appl. Environ. Microbiol.*, vol. 68, no. 5, pp. 2495-2502, May 2002.
- [5] S. Møller, C. Sternberg, J. B. Andersen, B. B. Christensen, J. L. Ramos, M. Givskov, and S. Molin "In situ gene expression in mixed-culture biofilms: evidence of metabolic interactions between community members," *Appl Environ Microbiol.*, vol. 64, no. 2, pp. 721-732, Feb. 1998.
- [6] S. Molin, and T. Tolker-Nielsen, "Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilization of the biofilm structure," *Curr. Opin. Biotechnol.*, vol. 14, no. 3, pp. 255-261, Jun. 2003.
- [7] L. H. G. Morton, D. L. A. Greenway, C. C. Gaylarde, and S. B. Surman, "Consideration of some implications of the resistance of biofilms to biocides," *Int. Biodeter. Biodegr.*, vol. 41, pp. 247-259, 1998.
- [8] J. W. Costerton, Z. Lewandowski, D. DeBeer, D. Caldwell, D. Korber, and G. James, "Biofilms, the customized microniche," *J. Bacteriol.*, vol. 176, no. 8, pp. 2137-2142, Apr. 1994.
- [9] G. A. O'Toole, and R. Kolter, "Initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis," *Mol. Microbiol.*, vol. 28, no. 3, pp. 449-461, May 1998.
- [10] G. A. O'Toole, H. B. Kaplan, and R. Kolter, "Biofilm formation as microbial development," *Annu. Rev. Microbiol.*, vol. 54, pp. 49-79, 2000.
- [11] K. Sauer, "The genomics and proteomics of biofilm formation," *Genome Biol.*, vol. 4, no. 6, pp. 219, May 2003.
- [12] S. Stammen, et al., "High-yield intra- and extra- cellular protein production using *Bacillus megaterium*," *Appl. Environ. Microbiol.* vol. 76, pp. 4037-4046, Jun. 2010.
- [13] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, "Gene regulation at the single-cell level," *Science*, vol. 307, no. 5717, pp. 1962-1965, Mar. 2005.
- [14] J. C. Locke, J. W. Young, M. Fontes, M. J. Hernández Jiménez, and M. B. Elowitz, "Stochastic Pulse Regulation in Bacterial Stress Response," *Science*, vol. 334, no. 6054, pp. 366-369, Oct. 2011.
- [15] M.B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain PS, "Stochastic Gene Expression in a Single Cell," *Science*, vol. 297, no. 5584, pp. 1183-1186, Aug. 2002.
- [16] P. T. Su, C. T. Liao, J. R. Roan, S. H. Wang, A. Chiou, and W. J. Syu, "Bacterial colony from two-dimensional division to three-dimensional development," *PLoS One*, vol. 7, no. 11, Nov. 2012.
- [17] A. Elfving, Y. LeMarc, J. Baranyi, and A. Ballagi, "Observing growth and division of large numbers of individual bacteria by Image Analysis," *Appl. Environ. Microbiol.*, vol. 70, no. 2, pp. 675-678, Feb. 2004.
- [18] Y. Wakamoto, J. Ramsden, and K. Yasuda, "Single-cell growth and division dynamics showing epigenetic correlations," *The Analyst*, vol. 130, no. 3, pp. 311-317, Mar. 2005.
- [19] D. Siegal-Gaskins and S. Crosson, "Tightly-regulated and heritable division control in single bacterial cells," *Biophys. J.*, vol. 95, no. 4, pp. 2063-2072, Nov. 2008.
- [20] T. Ross and T. A. McMeekin, "Modeling microbial growth within food safety risk assessments," *Risk Anal.*, vol. 23, no. 1, pp. 179-197, Feb. 2003.
- [21] J. M. Guberman, A. Fay, J. Dworkin, N. S. Wingreen, and Z. Gitai, "PSICIC: noise and asymmetry in bacterial division revealed by computational image analysis at subpixel resolution," *PLoS Comput. Biol.*, vol. 4, no. 11, Nov. 2008.
- [22] J. C. Locke, and M. B. Elowitz, "Using movies to analyse gene circuit dynamics in single cells," *Nat. Rev. Microbiol.*, vol. 7, no. 5, pp. 383-392, May 2009.
- [23] Q. Wang, J. Niemi, C. M. Tan, L. You, and M. West, "Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy," *Cytometry A.*, vol. 77, no. 1, pp. 101-110, Jan. 2010.
- [24] J. Klein, S. Leupold, I. Biegler, R. Biedendieck, R. Münch, and D. Jahn, "TLM-Tracker: software for cell segmentation, tracking and lineage analysis in time-lapse microscopy movies," *Bioinformatics*, vol. 28, no. 17, pp. 2276-2277, Sep. 2012.
- [25] O. Sliusarenko, J. Heinritz, T. Emonet, and C. Jacobs-Wagner, "High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics," *Mol. Microbiol.*, vol. 80, no. 3, pp. 612-627, May 2011.

- [26] M. N. Do, and M. Vetterli, "The contourlet transform: an efficient directional multi resolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091-2106, Dec. 2005.
- [27] M. N. Do, and M. Vetterli, "Contourlets," in *Beyond Wavelets*, 1st ed., G. V. Welland Ed., New York: Academic Press, 2003, pp. 83-105.
- [28] M. N. Do, and M. Vetterli, "Contourlets: a directional multiresolution image representation," in *IEEE International Conference on Image Processing Rochester*, 2002, pp. I-357 - I-360, vol.1.
- [29] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," *Graphic Gems IV*, San Diego: Academic Press Professional, 1994, pp. 474-485.
- [30] R. Gonzales and R. Woods, "Digital Image Processing," 3rd ed., Addison-Wesley Publishing Company, UK, 2007.
- [31] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 9, no. 1, pp. 62-66, Jan. 1979.
- [32] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986.
- [33] W. Margolin, "FtsZ and the division of prokaryotic cells and organelles," *Nature Rev. Mol. Cell. Biol.*, vol. 6, pp.862-871, Nov. 2005.
- [34] R. O. Duda, P. E. Hart, D. G. Stork, "Pattern Classification", 2nd ed., J. Wiley Ed., New York: John Wiley & Sons, 2001.
- [35] G. J. McLachlan and D. Peel, "Finite Mixture Models," 1st ed., J. Wiley Ed., New York: John Wiley & Sons, 2000.
- [36] G. J. McLachlan, T. Krishnan, "The EM Algorithm and Extensions," 2nd ed., J. Wiley Ed., New York: John Wiley & Sons, 2008.
- [37] S. Theodoridis and K. Koutroumbas, "Pattern Recognition," 4th ed., Elsevier, USA, 2008.
- [38] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no.1, pp. 113-125, Jul. 1994.
- [39] P. Tsakanikas, and E. S. Manolakos, "Protein spot detection and quantification in 2-DE gel images using machine-learning methods," *Proteomics*, vol. 11, no. 10, pp. 2038-2050, Apr. 2011.
- [40] M. Figueiredo, A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381-396, Mar. 2002.
- [41] K. P. Koutsoumanis and A. Lianou, "Stochasticity in colonial growth dynamics of individual bacterial cells," *Appl. Environ. Microbiol.*, vol. 79, no. 7, pp. 2294-2301, Apr. 2013.
- [42] K. Wang, L. Kang, A. Anand, G. Lazarovits, and K. S. Mysore, "Monitoring in plant a bacterial infection at both cellular and whole-plant levels using the green fluorescent protein variant GFPuv," *New Phytol.*, vol. 174, no. 1, pp. 212-23, 2007.
- [43] J. Baranyi and T. A. Roberts, "A dynamic approach to predicting bacterial growth in food," *Int. J. Food Microbiol.*, vol. 23, no. 3-4, pp. 277-294, Nov. 1994.
- [44] J. Lu, D. A. Carter, L. Turnbull, D. Rosendale, D. Hedderley, et al., "The Effect of New Zealand Kanuka, Manuka and Clover Honeys on Bacterial Growth Dynamics and Cellular Morphology Varies According to the Species," *PLoS ONE*, vol. 8, no. 2, Feb. 2013.
- [45] C.N. Takacs, J. Hocking, M.T. Cabeen, N.K. Bui, S. Poggio, et al., "Growth Medium-Dependent Glycine Incorporation into the Peptidoglycan of *Caulobacter crescentus*," *PLoS ONE*, vol. 8, no. 2, Feb. 2013.
- [46] J.B. Fiche, D.I. Cattoni, N. Diekmann, J.M. Langerak, C. Clerte, et al., "Recruitment, Assembly, and Molecular Architecture of the SpoIIIE DNA Pump Revealed by Superresolution Microscopy," *PLoS Biol.*, vol. 11, no. 5, May 2013.
- [47] D. Patzelt, H.Wang, I.Buchholz, M. Rohde, L. Gröbe, S. Pradella, A. Neumann, S. Schulz, S. Heyber, K. Münch, R. Münch, D. Jahn, I. Wagner-Döbler, and J. Tomasch, "You are what you talk: quorum sensing induces individual morphologies and cell division modes in *Dinoroseobactershibae*," *The ISME Journal*, Jul. 2013.
- [48] M. LeRoux, J. A. De Leon, N. J. Kuwada, A. B. Russell, D. Pinto-Santini, R. D. Hood, D. M. Agnello, S. M. Robertson, P. A. Wiggins, and J. D. Mougous, "Quantitative single-cell characterization of bacterial interactions reveals type VI secretion is a double-edged sword," *PNAS*, vol. 109, no. 48, pp. 19804-19809, Nov. 2012.
- [49] J. Lloyd-Price, A. Häkkinen, M. Kandhavelu, I. J. Marques, S. Chowdhury, E. Lihavainen, O. Yli-Harja, and A. S. Ribeiro, "Asymmetric disposal of individual protein aggregates in *Escherichia coli*, one aggregate at a time," *J. Bacteriol.*, vol. 194, no.7, pp. 1747-1752, Apr. 2012.
- [50] M. Kandhavelu, H. Mannerström, A. Gupta, A. Häkkinen, J. Lloyd-Price, O. Yli-Harja, and A. S. Ribeiro, "In vivo kinetics of transcription initiation of the *lar* promoter in *Escherichia coli*. Evidence for a sequential mechanism with two rate-limiting steps," *BMC Syst. Biol.*, vol. 5, no. 149, Sep. 2011.