

ADAPTIVE LOCALISATION AND TRACKING OF MULTIPLE MICROORGANISMS WITHIN A 4D HOLOGRAPHY DATASET



Marwan Elkholy
Department of Biomedical Engineering
McGill University
Montreal, Quebec, Canada
January 2017

A thesis submitted to McGill University in partial fulfillment of the requirements
of the degree of Master of Engineering

© Marwan Elkholy 2017



DEDICATION

To my mother Dr. Azza Aly, M.D.

The original microbiologist of our family,
who was not able to complete her Master's Thesis because of her difficult pregnancy with me.

Thank you for being a constant pillar of love and support.

ACKNOWLEDGEMENTS

I cannot express enough gratitude towards Professor Jay Nadeau for her constant guidance throughout my studies. She always made herself present and promptly responded to all my inquiries, whether research-related or otherwise, when I was struggling and confused. I am especially thankful that she has shown me the beauty of research by demonstrating her passion towards her work every day.

My completion of this project would not have been possible without the support of my lab-mates, Specifically Manuel Bedrossian whose endless curiosity inspired me to take on challenging endeavours; Dr. James Kent Wallace who has always been supportive and encouraging; and Dr. Christian Lindensmith who inspires a “nothing is impossible” attitude. Without you, my accomplishments over these past two years would not exist.

I would also like to acknowledge Dr. Iman Haji-Albolhassani, who sparked my interest in Machine Learning algorithms through our engaging discussions in the lab; and Pina Sorrini who made the department feel incredibly warm and welcoming.

Finally I would like to thank all my friends and family. Notably, Professor Ahmed Elkholy who has always pushed me to excel in everything I do; and Mohamed El Gindi who took an active interest in discussing my research, then helped to significantly optimise my work-flow.

I stand on the shoulders of giants.

Financial support for this project was provided by the Betty and Moore Foundation. Thank you for making this happen.

ABSTRACT

Understanding microorganisms' motility provides valuable insight into multiple fields including infectious disease, environmental microbiology, and astrobiology. However, observing microbial motility using conventional light or fluorescence microscopy allows for accurate localisation in two-dimensional (2D) space only, despite the fact that the natural microbial environment exists in three-dimensional (3D) space.

Digital Holographic Microscopy (DHM) has recently emerged as a powerful tool for studying microorganisms within their unconstrained 3D space. Unfortunately, this technique suffers from a poor signal-to-noise ratio (SNR), especially when using unlabeled cells (without fluorescent proteins or dyes). Therefore, automatically tracking the bacterial trajectories using image processing and computer vision methods is challenging.

This thesis presents a supervised machine-learning classifier that can handle the low SNR DHM phase image reconstructions, and accurately localise unlabeled cells' centroids with an error margin of approximately a cell length in 3D space. The classifier was validated against two manually labelled gold standard datasets, each containing a specific species of bacteria (*Bacillus subtilis* and *Colwellia psychrerythraea*). This classifier achieved a recall of over 50% and a precision of over 90% in both validation datasets. In conclusion, machine-learning classifiers are easy to implement and can successfully extract useful information from noisy DHM datasets.

RÉSUMÉ

La motilité des microorganismes fournit des informations précieuses sur beaucoup de domaines, y compris les maladies infectieuses, la microbiologie environnementale et l'astrobiologie. Cependant, l'observation de la motilité microbienne à l'aide d'une microscopie conventionnelle de lumière ou de fluorescence permet une localisation précise dans l'espace bidimensionnel (2D) seulement, malgré le fait que l'environnement microbien naturel existe dans l'espace tridimensionnel (3D).

La microscopie holographique numérique (DHM) est récemment apparue comme un outil puissant pour étudier les microorganismes dans leur espace 3D non contraint. Malheureusement, cette technique souffre d'un mauvais rapport signal / bruit (SNR), en particulier lors de l'utilisation de cellules naturelles (sans protéines fluorescentes ni colorants). Par conséquent, le suivi automatique des trajectoires bactériennes à l'aide de méthodes de traitement d'image et de vision par ordinateur est difficile.

Cette thèse présente un classificateur machine-learning supervisé qui peut traiter les reconstructions d'image de phase de DHM avec SNR faible, et localiser les centroïdes des cellules naturelles dans une marge d'erreur d'approximativement une longueur de cellule dans l'espace 3D. Le classificateur a été validé contre deux ensembles de données étiquetés manuellement, où chaque ensemble de données contient une seule espèce de bactéries (*Bacillus subtilis* ou *Colwellia psychrerythraea*). Ce classificateur a obtenu un rappel de plus de 50% et une précision de plus de 90% dans les deux ensembles de données de validation. En conclusion, les classificateurs machine-learning sont faciles à mettre en œuvre et réussissent à extraire des informations utiles à partir d'ensembles de données DHM bruyants.

LIST OF FIGURES

Figure 1: Typical “random walk” path taken by <i>E. coli</i>	13
Figure 2: Bacterial trajectories of <i>Vibrio alginolyticus</i>	16
Figure 3: Solid sea ice under microscopic inspection.	18
Figure 4: Comparison of 3D and 2D tracking methods [17].	20
Figure 5: Reconstructed 3D trajectory for a bacterium.....	21
Figure 6: Michelson interferometer for digital holography.....	24
Figure 7: Mach-Zehnder interferometer for digital holography	24
Figure 8: Basic configuration a the Mach-Zender DHM	27
Figure 9: Examples of simulated image data.	30
Figure 10: Schematic of the DHM used in this thesis	34
Figure 11: Field implementation of th DHM used in this thesis	35
Figure 12: Sample chamber design schematic used in this thesis.....	36
Figure 13: 3D tracks of <i>Vibrio alginolyticus</i>	38
Figure 14: Algorithm Flowchart	39
Figure 15: Object Identification Methodology Overview	40
Figure 16: Image classification results.	40
Figure 17: Linear Logistic Regression Classifier function	43
Figure 18: Pre-processing results.....	46
Figure 19: User-selected cell’s pixels	47
Figure 20: Labelling a new group of pixels as bacteria	50
Figure 21: The 8 Microbial <i>B. subtilis</i> Tracks within Validation Dataset 1.....	56
Figure 22: The 6 <i>Colwellia psychrerythraea</i> tracks within Validation Dataset 2	57
Figure 23: Post-Classification Binary Image within Validation Dataset 1.....	60
Figure 24: Post-Classification Binary Image within Validation Dataset 2	60
Figure 25: automatically identified points within Validation Dataset 1	62
Figure 26: Top view of automatically labelled points in Validation Dataset 1	62
Figure 27: Histogram of error distances in Validation Dataset 1	64
Figure 28: automatically identified points within Validation Dataset 2	65
Figure 29: Top view of automatically labelled points in Validation Dataset 2	66
Figure 30: Histogram of error distances in Validation Dataset 2	67
Figure 31: 3D plot of automatically identified tracks in Validation Dataset 1	69
Figure 32: top view of automatically labelled tracks in Validation dataset 1	69
Figure 33: 3D plot of automatically labelled tracks in Validation Dataset 2	70
Figure 34: Top view of automatically labelled tracks in Validation Dataset 2	70
Figure 35: Instantaneous motility direction of the tracks in Validation Dataset 1	72
Figure 36: Instantaneous motility direction of the tracks in Validation Dataset 2	73
Figure 37(a, b): Histogram of microorganism speeds in Validation Dataset 1	75
Figure 38: Manually Labelled 2D speed in the Validation Dataset 1	75

LIST OF TABLES

Table 1 Properties of the simulated image data used to evaluate tracking methods	30
Table 2 The 14 methods used in the comparative study to detect and link particles	31
Table 3 Properties of the DHM used in this thesis.	35
Table 4 Generalised input and output of the algorithm described in this thesis	38
Table 5 Variable and Parameter list of the object identification algorithm.....	41
Table 6 Pixel Features used in the classifier	44
Table 7 Variables and Parameters list of the object tracking algorithm	52
Table 8 Properties of the Gold Standard Datasets used for validation	55
Table 9 Parameter values assigned using Training dataset of Validation Dataset 1.	61
Table 10 Validation metrics of the algorithm using Validation Dataset 1	63
Table 11 Parameter values assigned using Training dataset of Validation Dataset 2. ...	64
Table 12 Validation metrics of the algorithm using Validation Dataset 2	66

LIST OF ABBREVIATIONS

2D, 3D	Two-dimensional, three-dimensional
CCD	Charge-coupled device
DHM	Digital Holographic Microscope
EDTA	Ethylenediaminetetraacetic acid
FN	False Negatives
FP	False Positives
fps	Frames-per-second
GPU	Graphical Processing Unit
HI	Holographic Interferometry
Hz	Hertz
LB	Lysogeny Broth
LR	Logistic Regression
mm	Millimeters
MO	Microscope Objective Lens
Mpx	Megapixels
nm	Nanometers
rad	Radians
RMS	Root-mean-squared
SNR	Signal-to-noise ratio
TP	True Positives

TABLE OF CONTENTS

Dedication	2
Acknowledgements	3
Abstract	4
Résumé.....	5
List of Figures	6
List of Tables	7
List of Abbreviations	8
Chapter 1: Introduction	11
1.1 Project Summary	11
1.2 Motivation	11
1.3 Justification	12
1.4 Contributions	12
Chapter 2: Background and Literature review.....	13
2.1 Why Study Microbial Motility?	13
Understanding Microbial Biophysical Mechanisms	13
Motility as a tool for Astrobiology.....	17
2.2 Traditional Methods of Studying Microbial Motility	19
Why use 3D measurements instead of 2D?	19
Methods for observing microorganisms in 3D	20
2.3 Digital Holographic Microscopy	23
Holographic interferometry	23
Reconstructing a Hologram	25
Microscopy using Holographic interferometry	26
Biological Challenges of using a Digital Holographic Microscope	27
2.4 Tracking Algorithms for Observing Microbial Motility	29
Chapter 3: Methods and Experimental Techniques	34
3.1 Data Acquisition	34
Specifications and Design of the Digital Holographic Microscope.....	34

Sample Chamber.....	36
Preparation of Microbial Samples.....	36
Reconstruction of a Hologram into multiple 2D z-planes.....	37
3.2 Generating Motility tracks.....	38
Overview.....	38
3.2.1 - Object Identification.....	40
3.2.2 - Generating Tracks.....	52
3.2.3 - Validation.....	54
3.3 Optimizing Computational Speed.....	58
Optimizing Ram Usage.....	58
Using a GPU for image pixel operations.....	58
Chapter 4: Results and Observations.....	59
Overview.....	59
4.1 Particle Identification Algorithm Results.....	60
Particle Identification Results from Validation Dataset 1.....	61
Particle Identification Results from Validation Dataset 2.....	64
4.2 Particle Tracking Results.....	68
4.3 Properties of Microbial Motility.....	71
Swimming Direction.....	71
Swimming Speed.....	74
Chapter 5: Discussion, Future Work and Conclusions.....	76
5.1 Discussion.....	76
5.2 Future Work.....	78
Investigate the Time Stamps.....	78
Investigate different 3D features of the Algorithm.....	79
Investigate the Effect of Microbial Density on Algorithm Performance.....	80
Create a Motility Database.....	80
5.3 Conclusion.....	81
References.....	82

CHAPTER 1: INTRODUCTION

1.1 Project Summary

This thesis describes a machine learning algorithm that is used to obtain three-dimensional trajectories of live, lab-grown microorganisms as they move within an essentially unrestrained environment. The microorganisms are imaged using a Digital Holographic Microscope (DHM), which captures three dimensional holograms of the entire sample chamber at a rate of about 10Hz. Thanks to the phase contrast imaging capability of DHM, the microorganisms do not need to be labeled; all contrast enhancing techniques are performed numerically in post-processing after the hologram is captured.

1.2 Motivation

Current techniques for observing bacterial motility are essentially two-dimensional because of the small depth of field provided by high numerical aperture objectives. Measurement of 3D trajectories is performed by approximating the third dimension from measured 2D trajectories, or by inferring the organisms' z positions as they travel into and out of focus. The boundaries of the coverslips constrain motion in the z direction and affect the hydrodynamics of the motility and the organisms' possible swimming ranges. This means that for most bacteria, their swimming patterns in an unconstrained 3D volume remain essentially unknown. Studying the 3D trajectories of microorganisms is important in numerous fields, including infectious disease, environmental microbiology, and astrobiology. See section 2.1 for more information.

1.3 Justification

Digital Holographic Microscopy has recently emerged as a powerful tool for studying microorganisms within their unconstrained 3D space. Another important tool that has recently emerged is machine learning algorithms that find trends in data without having to be explicitly programmed. Both of these tools have become available thanks to the proliferation of computational power along with an abundance of computational algorithms. This thesis exploits these advances by presenting a simple yet accurate way of locating objects of interest within a large, noisy 3D dataset, and tracking all the objects simultaneously with high throughput.

1.4 Contributions

Logistic regression (LR) is the machine learning pixel classification tool used to achieve the results of this study. Although LR has been previously used for pixel classification of images [1, 2], it has never been applied to 3D holography reconstructions as a tool for object detection. This study presents LR being used for localisation of different microorganism species in 3D, demonstrating the flexibility of the classifier and its ability to 'learn' the properties of the microorganisms of interest.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

2.1 Why Study Microbial Motility?

Microbial processes are often governed by the microorganisms' motility strategies. Examples of microbial processes include behavioural responses of microbial cells to chemical stimuli, the interactions between cells and biotic or abiotic surfaces, and cell–cell interactions in microbial suspensions. Understanding the microbes' motility strategies provide valuable insight into a broad range of fields such as infectious disease, environmental microbiology, and astrobiology. Figure 1 shows a typical random walk path taken by *E. coli* [3]. It demonstrates the “run” and “tumble” motility strategies visible through a traditional light microscope.

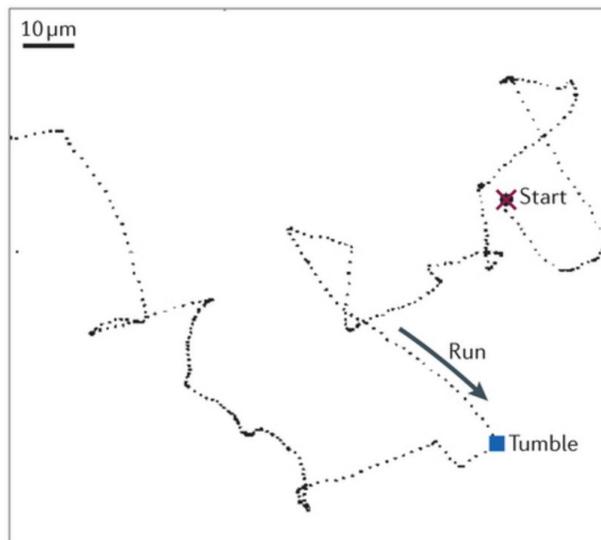


Figure 1

Typical “random walk” path taken by *E. coli* and other bacteria with multiple flagella. This figure demonstrates the “run” and “tumble” motility strategies. [3]

UNDERSTANDING MICROBIAL BIOPHYSICAL MECHANISMS

Motility models are constantly being proposed and verified using the latest technologies available. For example back in 1973, Berg et al. [4] suggested that bacteria swim by rotating their

flagellar filaments. Using mathematical models, they found that the power required for propulsion was consistent with laboratory observation, given that multiple flagella rotate together in bundles to propagate helical waves. In 1974, Silverman et al. [5] used light microscopy to observe the motility of *E. coli* mutants. The mutants lacked genes used to construct parts of the flagellar filament and the flagellar hook structure. Silverman et al. then obtained electron micrographs of the flagellar structures, and inferred that the flagellar hook structure is driven in a rotary fashion, and results in the rotation of the flagellar filament. Further confirmation that bacteria swim by rotating their flagellar filaments was obtained in 1977, when Macnab et al. [6] investigated flagellar rotation by constructing physical models of the entire flagellar structures made from stainless steel and Teflon, with parameter properties matching those typically observed by light and electron microscopy. These experiments provided the foundation of bacteria propulsion biophysics.

In more recent years, researchers attempted to explain how bacteria were able to direct their propulsion specifically towards chemical attractants, and away from chemical repellants. In 2000, Turner et al. [7] used advances in fluorescence microscopy to perform real time imaging of *E. coli*'s flagellar filaments. They confirmed that *E. coli* cells modulated the direction of rotation, from counter-clockwise to clockwise, of one or many of their flagellar motors to change their swimming direction. This is called a "tumble". Figure 1 displays a typical *E. coli* random walk observable by a microscope, demonstrating the "run" and "tumble". A run is when the bacterium moves in a direct, straight direction. A tumble is when the bacterium changes its orientation at the end of a run, before performing another run in a new direction. Their study was limited to the 167 times that a cell ran into the field of view within the plane of focus, tumbled within the

plane of focus, and ran out of the field of view still within the plane of focus of the fluorescence microscope. These events were relatively rare, but they allowed the researchers to analyze them in their entirety.

Other bacterial genera, such as *Vibrio*, possess a single flagellum instead of the multiple flagella that permit *E. coli* to run and tumble. The single flagellum has a bidirectional motor similar to *E. coli*, giving these species two motility strategies: “forward” (flagellum runs one way) and “reverse” (flagellum runs the other way). Early studies of these organisms, such as one by Taylor et al. [8], had proposed that these species change direction because each reversal typically results in a small change in cell orientation, allowing the bacterium to significantly change its swimming direction over multiple reversals. In 2005, Magariyama et al. [9] used 2D phase-contrast microscopy to observe the swimming patterns of three different strains of *Vibrio alginolyticus*--one with only forward swimming, one with only backwards swimming, and the wild-type bacteria with both forwards and backwards swimming. Using only the microorganisms that were within the narrow depth of focus of their microscope (~10 μm), they observed that near the surface, the forward direction (trajectory?) was straight and the backwards direction was significantly curved, forming a tight circle a few bacterial lengths in diameter. The asymmetry in swimming direction was explained by a hydrodynamic interaction with the surface [10].

In 2011, Xie et al [11] proposed a new motility strategy for bacterial species with a single polar flagellum: the three-step (forward-reverse-flick) swimming pattern for chemotaxis away from the surface. The researchers obtained their data using an inverted microscope coupled with a 30 fps CCD camera. Only the bacteria present within the narrow depth of focus of their

microscope ($\sim 6 \mu\text{m}$) that demonstrated two flicking events were kept, and the rest were discarded. Figure 2 shows a typical swimming path taken by *Vibrio alginolyticus*.

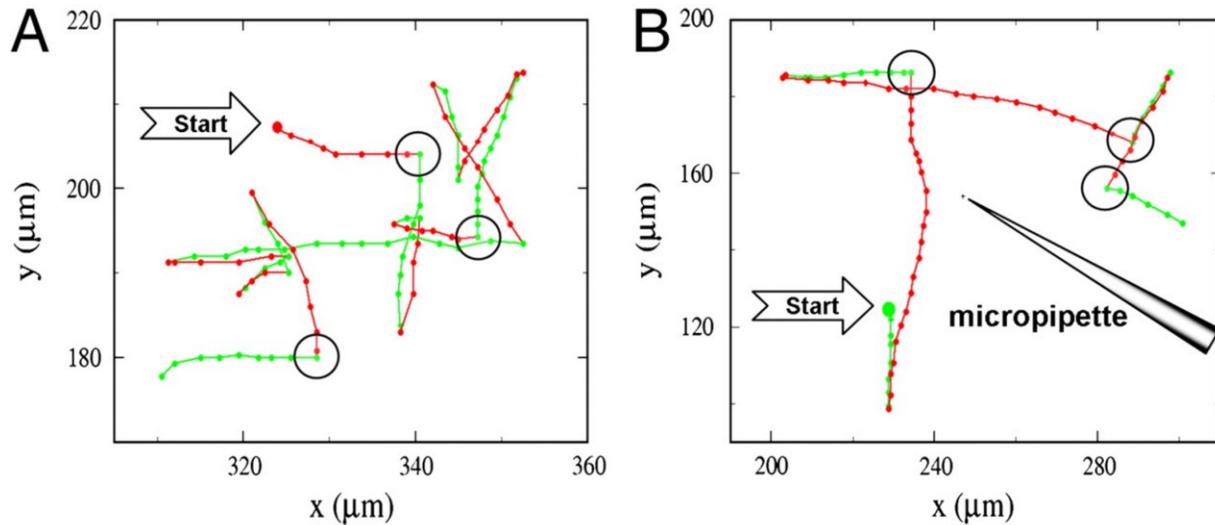


Figure 2

Bacterial trajectories of *Vibrio alginolyticus* in a motility medium (A) and in a steep chemical gradient created by a micropipette filled with 1 mM serine (B). The big solid circles are the starting points of the bacterial tracks, and the small solid circles represent the positions at an equal time interval of 0.067 s. The green and the red segments correspond to the forward and the backward trajectories, respectively. The large open circles marked the flicking events; for clarity, not all flicking events are marked in A.

[11]

The studies described in this section shed significant light onto understanding microbial biophysical mechanisms. However, they were all performed as a 2D analysis of microbial motility patterns, and are therefore limited by the narrow depth of focus of most conventional microscopes. The world we live in exists in three dimensions, and therefore critical information may have been missed or lost due to the limitations of the technologies available to the researchers. New technologies such as Digital Holographic Microscopy (DHM) have emerged to give us further insight into 3D microbial motility. DHM will be discussed in detail in sections 2.3.

MOTILITY AS A TOOL FOR ASTROBIOLOGY

In the 17th century, a lens maker named Antonie Van Leewenhoek observed living microorganisms using a single-lens microscope [12]. The motility he observed gave him immense delight when he proclaimed: “I must say, for my part, that no more pleasant sight has ever yet come before my eye than these many thousands of living creatures, seen all alive in a little drop of water, moving among one another, each several creature having its own proper motion.” [13]. This marked the beginning of the field of microbiology since, prior to his observation, the existence of sub-visible microorganisms was not widely accepted.

These events demonstrate the utility of characteristic microbial motility as a biosignature—that is, as a means for recognizing extraterrestrial microbial life as life even if it is physically and chemically very different from Earth life. Nadeau et al. [14] proposed motility as the most unambiguous biosignature for life detection missions within the Solar System, where liquid samples may be collected and imaged *in situ*. They argued that microbial motility on Earth was ubiquitous, even in extremely cold extraterrestrial analogue environments such as sea ice, glacier ice, and permafrost. The study presented images of brines collected from sea ice. The ice appears solid to the naked eye, but microscopic inspection reveals that it is in fact porous, containing a network of microscopic veins and channels filled with brine (see Figure 3), which may be collected as it oozes slowly from a cut surface of sea ice at sub-zero ambient temperatures. The researchers found microorganisms, including prokaryotes, living within these veins and channels by simply observing the motility of bacteria and algae. The organisms were unambiguously identified as alive and active because directed motion by living microorganisms

is distinct from diffusion and flow of dead matter. While not all bacteria present were motile, all observed samples contained some motile organisms.

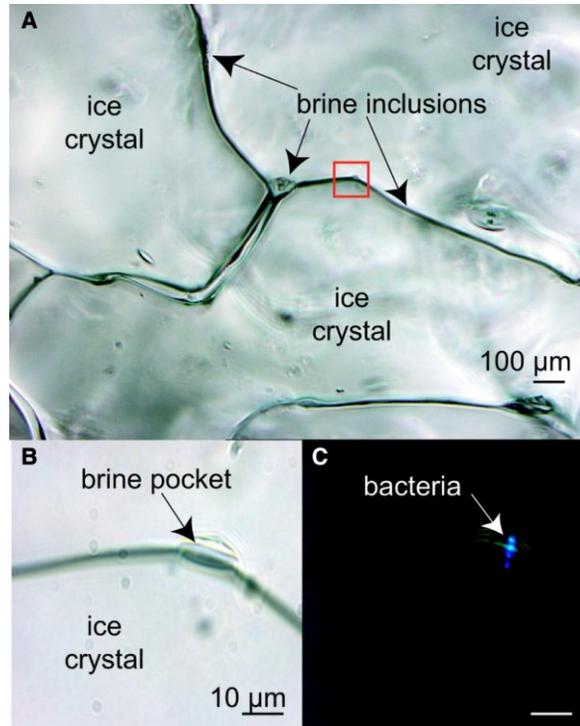


Figure 3
Solid sea ice under microscopic inspection. The area outlined in (A) is enlarged in (B) and (C), where (C) was taken by epifluorescence microscopy following staining with the DNA-specific stain DAPI.

[14]

2.2 Traditional Methods of Studying Microbial Motility

WHY USE 3D MEASUREMENTS INSTEAD OF 2D?

Observing motility using conventional light microscopy only allows for localisation in 2D. However, the natural microbial environment exists in 3D. Conventional observations may either capture a projection of the 3D trajectories onto the 2D plane [15], or they may be limited to only the z-planes that are in focus [7, 9, 11, 16]. This severe limitation gives an incomplete image of the motility patterns observed. For example, a bacterium travelling up, perpendicular to the plane of projection, will appear stationary using conventional techniques. Taute et al. [17] calculated the systematic errors that arise using conventional microscopic techniques when observing bacterial motility (see Figure 4). They found that aside from the effects of localisation errors, 2D projection (column 2 of Figure 4) of the same volume introduces systematic errors in speed and turning angle measurements, compared to the correct speed and turning angle measurements found in 3D tracking (column 1 in Figure 4). On the other hand, observations obtained from 2D slicing (column 3 of Figure 4) are constrained to a thin focal plane of thickness d . Assuming that runs have to lie fully within the slice and are five times longer than d , the vast majority of turning events are ignored and a bias against turning angles near 90° is introduced. In summary, 2D methods do not capture the entire complexity of bacterial motility within their natural, 3D environments.

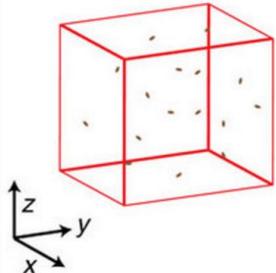
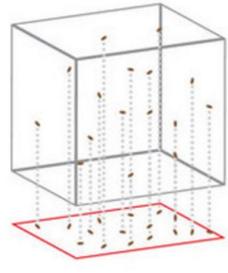
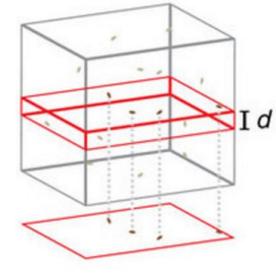
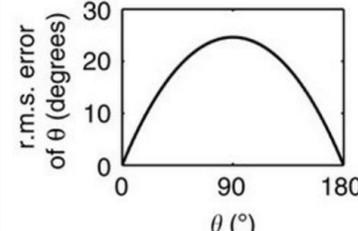
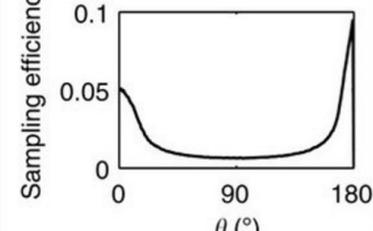
	Column 1	Column 2	Column 3
Method	<p>3D</p> 	<p>2D projection</p> 	<p>2D slicing</p> 
Speed	Correct	>30% r.m.s. error	Error ≤ 2%
Turning angles	Correct		

Figure 4
Comparison of 3D and 2D tracking methods [17].

METHODS FOR OBSERVING MICROORGANISMS IN 3D

There are few techniques that allow for capturing bacterial motility in 3D. Howard C. Berg [18] managed to manually track bacteria by moving the sample in the plane of focus to keep a bacterium in focus back in 1971. This has since been the gold standard of bacterial 3D tracking. However this method is technically demanding because it requires a specialised experimental setup, and only allows for tracking one bacterium at a time. A simpler way to measure depth is to use the largest diffraction ring diameter of the object being observed as a linear measure of the object's absolute distance $|z|$ from the focal plane [19]. This method has high throughput,

and has been done with fluorescence microscopy [20], darkfield microscopy [21] and phase contrast microscopy [22]. However, these methods only allow for a maximum depth measurement of about $\pm 30\mu\text{m}$ from the focal plane before compromising the accuracy of the measurement. Taute et al. [17] improved upon these methods to get a larger range of $200\mu\text{m}$ of depth measurements using phase contrast microscopy. They achieved this by using image cross-correlations of the observed bacterial diffraction patterns with a “diffraction patterns reference library” to infer the z-position of the bacteria (see Figure 5). However their method is most suitable to tracking spherical bacteria of a specific size, because the method requires the creation of a reference library using fixed sized silica beads.

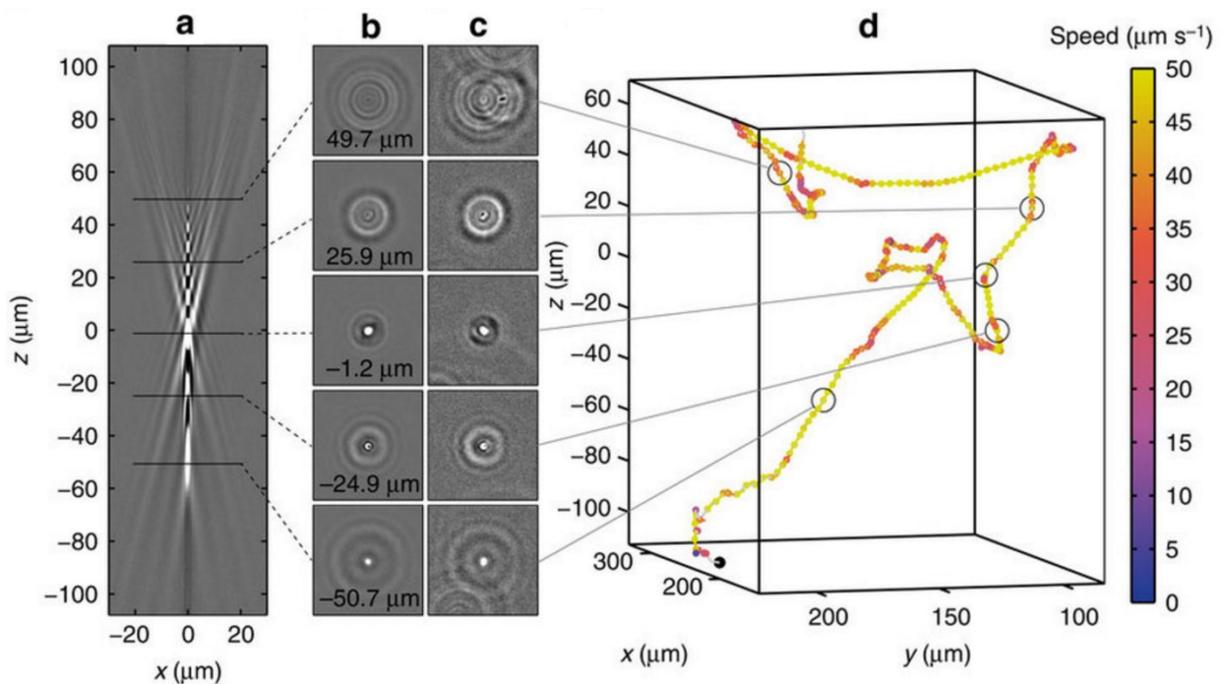


Figure 5

(a) A vertical slice through a reference library created by combining 73 aligned image stacks obtained for $1\mu\text{m}$ silica beads.

(b) Horizontal slices from the reference library at positions marked in a.

(c) Images of a swimming *E. coli* bacterium at the corresponding positions.

(d) Reconstructed 3D trajectory for the bacterium in c. The trajectory starting point is marked by a black dot.

[17]

Digital Holographic Microscopy (DHM) has emerged as the most suitable method for measuring 3D motility information of live microbial cells. It overcomes the two main limitations discussed previously in this section: the need for a dynamic stage that must be adjusted to keep a bacterium in focus, and the need for a reference library to infer the position of regularly shaped bacteria from their out of focus diffraction patterns. DHM will be discussed in detail in section 2.3.

2.3 Digital Holographic Microscopy

Digital Holographic Microscopy (DHM) is based on the technique of holographic interferometry (HI). This section will overview HI, and describe how HI is used along with a magnifying objective (MO) lens to create a holographic microscope.

HOLOGRAPHIC INTERFEROMETRY

In 1948, Dennis Gabor [23, 24] invented holography as a technique for recording the amplitude and phase of a wavefield. The recorded interference pattern between an object's scattered wavefield and a coherent background, called the reference wave, is called a hologram [25]. Holograms are recorded as a flat 2D image, but their interference patterns contain information about the entire three-dimensional wavefield. The original object wave can be reconstructed from its hologram by illuminating the hologram with its same reference wave again. This reconstructed wave is indistinguishable from the original object wave, and an observer would see a 3D image. The word *holography* is derived from the Greek word 'holos' meaning 'whole' or 'entire', and 'graphein' meaning 'to write'.

An important application of holography is holographic interferometry (HI). HI was developed in 1965 by Stenson and Powell [26, 27], allowing researchers to map the 3D displacements of rough surfaces with sub-micrometer accuracy. There are two main types of interferometers--the Michelson interferometer for reflective objects and the Mach-Zehnder interferometer for transmissive objects [28]. A typical Michelson interferometer is presented in Figure 6, whereas a typical Mach-Zehnder interferometer is presented in Figure 7. In both diagrams, the light-green beams are the input from the laser, the light blue is the reference beam path, and the light red

depicts image formation of an object point. In both designs, the object is illuminated with a plane wave, and the reference arrives at the charged coupled device array (CCD) plane with the same wavefront curvature as the object wave. The CCD is a light sensor, used to digitally record the resulting hologram. The digital hologram can be reconstructed back into the original object wavefront using numerical methods implemented by a computer [29].

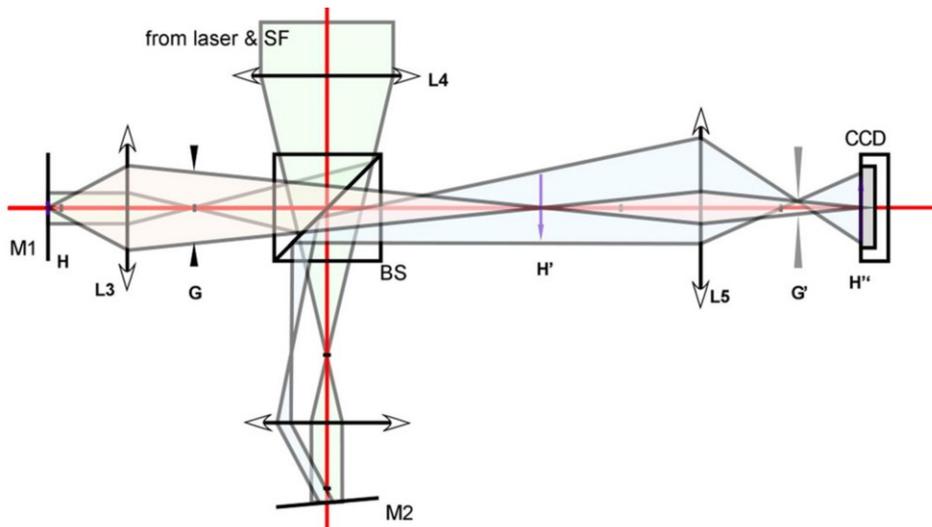


Figure 6
Michelson interferometer for digital holography of reflective samples.
BS: beamsplitters; L: lenses; H: image; M: mirrors; G: apertures.
[28]

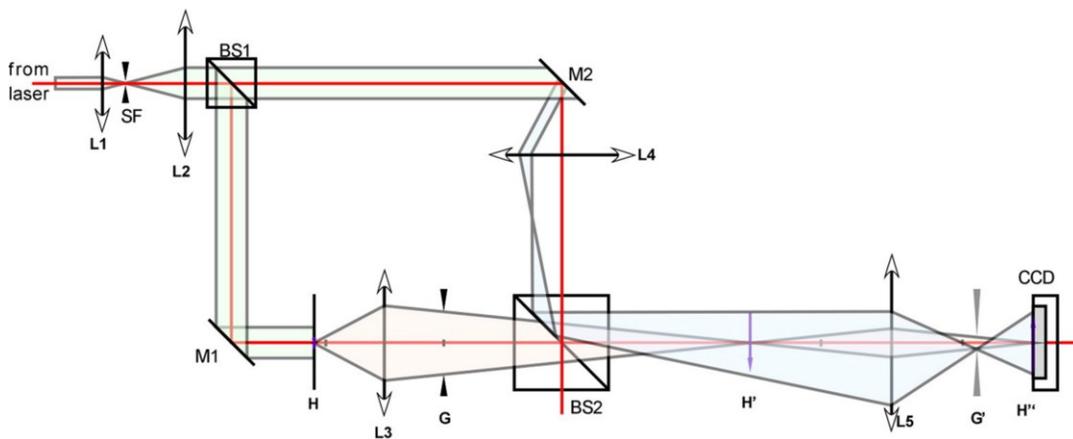


Figure 7
Mach-Zehnder interferometer for digital holography of transmissive samples
BS: beamsplitters; L: lenses; H: image; M: mirrors; G: apertures.
[28]

RECONSTRUCTING A HOLOGRAM

The amplitude and phase distribution in the plane of the real image can be found by the Fresnel–Kirchhoff integral [29, 30]. If a plane wave illuminates the hologram located in the plane $z = 0$, with an amplitude transmittance $t(x, y)$, the Fresnel–Kirchhoff integral gives the complex amplitude $\Gamma(\xi, \eta)$ in the plane of the real image. The amplitude and phase distribution can be approximated by the Fresnel approximation:

$$\begin{aligned} \Gamma(\xi, \eta) = & \frac{ia}{\lambda d} \exp \left[-i \frac{\pi}{\lambda d} (\xi^2 + \eta^2) \right] \\ & \times \iint_{(x,y)} t(x, y) \exp \left[-i \frac{\pi}{\lambda d} (x^2 + y^2) \right] \\ & \times \exp \left[+i \frac{2\pi}{\lambda d} (x\xi + y\eta) \right] dx dy. \end{aligned} \quad (2.1)$$

Parameter d is the distance between the object and the CCD array, a is the amplitude of the incident wave, and λ is the wavelength of the incident wave. The Fresnel approximation is valid if d is large enough, specifically:

$$d^3 \gg \frac{\pi}{4\lambda} [(\xi - x)^2 + (\eta - y)^2]^2. \quad (2.2)$$

The maximum possible value of $(\xi - x)^2$ and $(\eta - y)^2$ must be considered. For example, if $\lambda = 600$ nm and typical hologram dimensions of $(\xi - x)_{\max} = (\eta - y)_{\max} = 0.5$ cm, d must be much larger than 15 cm.

The intensity I in the real image can be calculated by squaring the modulus:

$$I = \text{Re}(\Gamma)^2 + \text{Im}(\Gamma)^2 \quad (2.3)$$

The phase image Φ can be obtained by the argument:

$$(2.4)$$

$$\Phi = \arctan \frac{\text{Im}(\Gamma)}{\text{Re}(\Gamma)}$$

The function $\Gamma(\xi, \eta)$ can be digitized if the hologram transmission $t(x, y)$ is sampled on a rectangular raster of $N \times N$ matrix points, with steps Δx and Δy along the coordinates. ξ and η are replaced by $r\Delta\xi$ and $s\Delta\eta$, where r and s are integers. In this case the discrete representation of the Fresnel approximation is given by the following equation [31]:

$$\begin{aligned} \Gamma(r, s) = & \exp \left[-i \frac{\pi}{\lambda d} (r^2 \Delta\xi^2 + s^2 \Delta\eta^2) \right] \\ & \times \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} t(k, l) \exp \left[-i \frac{\pi}{\lambda d} (k^2 \Delta x^2 + l^2 \Delta y^2) \right] \\ & \times \exp \left[i 2\pi \left(\frac{kr}{N} + \frac{ls}{N} \right) \right]. \end{aligned} \quad (2.5)$$

$\Gamma(r, s)$ is a matrix of $N \times N$ points that describes the amplitude and phase distribution of the real image. $\Delta\xi$ and $\Delta\eta$ are the pixel sizes in the reconstructed image. In summary, this equation is a representation of the Fresnel approximation in terms of the discrete Fourier transform used to digitally reconstruct a hologram image back to its original waveform, and ready for processing.

MICROSCOPY USING HOLOGRAPHIC INTERFEROMETRY

The ability of digital holographic interferometry to numerically focus on different sample planes without any optomechanical movement has been exploited to give rise to Digital Holographic Microscopy (DHM). A microscope objective (MO) enlarges the light transmitted by the sample (S) to form the object wave **O**, which interferes with the reference wave **R** to create the hologram. Figure 8 shows the schematic of a standard Mach-Zender DHM used by Marquet et al. [32] to study living neurons in 3D space. One can easily observe that a DHM is nothing more than an interferometer containing a MO within the beam paths.

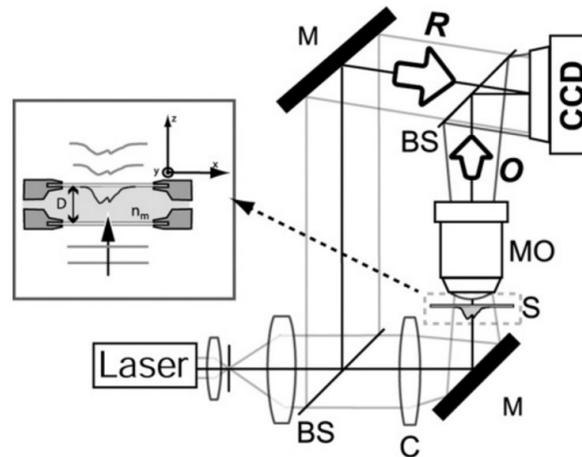


Figure 8

Reprinted with permission from ref [32], Optics Letters:
Basic configuration of the Mach-Zender DHM used by Marquet et al. [32]. Inset, schematic representation of cultured cells mounted in a closed perfusion chamber. M, mirror. BS, beam splitter. C, Condenser. n_m , refractive index of medium. D, thickness of sample chamber. Other abbreviations defined in text.

DHM has proven to be a powerful tool by enabling the study of biological entities in 3D space. It has been used to study distribution and swimming patterns of plankton in the open ocean [33], to investigate dinoflagellate feeding behaviour [34, 35], to study the motility of algal zoospores [36], and to study cultured cells in the laboratory [32, 37-42]. The CCD has the ability to digitally record many holograms per second and hence, capture the dynamic movements of microorganisms in 3D. In conclusion DHM is a powerful tool that is starting to gain traction within the life sciences community.

BIOLOGICAL CHALLENGES OF USING A DIGITAL HOLOGRAPHIC MICROSCOPE

There are challenges in biology that are not encountered in the more common material science applications of DHM. Specifically, cells are not perfect spheres, vary in size and shape from one cell to another, and have indices of refraction close to that of water [43]. The power of the light source must also be monitored as to not harm the organisms being observed, which

results in errors due to an increase in shot noise that becomes significant at low illumination [44]. Other sources of noise specific to DHM include laser speckle [45], which is inherit to any imaging technique using coherent light sources, and temporal phase noise, which results from the uncorrelated noise between the two beams of the interferometer.

These all result in difficulties with hologram reconstruction and data analysis by generating a low signal-to-noise ratio of the imaged microorganisms. Therefore, automatically identifying the microorganisms using standard image processing and computer vision methods is very challenging and requires the development of a sophisticated technique that can account for these noisy datasets. Furthermore, tracking the trajectories of each microorganism in 3D using DHM becomes much more complicated if the microorganisms cannot be effectively identified. This thesis solves the problem of bacterial identification in noisy datasets by using a machine learning algorithm that adapts to each dataset. It is the first time that an automated algorithm has been shown to work for holographic tracking of bacteria. Previous attempts to track bacteria with DHM have required complex de-noising algorithms coupled with manual tracking [46].

2.4 Tracking Algorithms for Observing Microbial Motility

Particle tracking has been used to quantitatively study many different dynamic biological processes such as cell membrane dynamics [47], cytoskeletal filaments, focal adhesions [48], viral infection [49], intracellular transport [50], gene transcription [51] and genome maintenance [52]. Technological developments in recent decades, such as the abundance of computational power along with the proliferation of computer algorithms, have resulted in the emergence of numerous automated particle tracking tools [53]. Automated particle tracking software tools can generally be divided into two steps: particle identification/detection (the spatial aspect), followed by particle tracking/linking (the temporal aspect). In the first step, the coordinates of particles of interest are found within every frame of the image sequence. In the second step, detected particles are connected from frame to frame to form tracks. Many algorithms have been developed over the years for each of these steps; some are specific to bioimaging [54-59], while others are general and can be applied to any given data set [60].

In 2014, Chenouard et al. [61] provided an objective comparative study of the most common particle tracking methods used in bioimaging. First, they identified three main factors that affect tracking performance: dynamics (type of motion), density (number of particles per field of view), and signal relative to noise (SNR). Second, they simulated a set of 2D and 3D image data based on these different factors (see Table 1 and Figure 9). They then sent these image data sets to 14 teams who took up the challenge of identifying and tracking the particles, using already established state-of-the-art methods (see Table 2). The teams then sent back their results, and their performance was quantitatively evaluated.

Table 1
Properties of the simulated image data used by Chenouard et al. to evaluate different particle tracking methods [61]

Parameter	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Nickname	Vesicles	Microtubules	Receptors	Viruses
Dynamics	Brownian	Directed	Switching	Switching
PSF model	Widefield	Non-isotropic Gaussian	Confocal	Confocal
Dimensionality	2D+time	2D+time	2D+time	3D+time
Image size (pixels)	512 × 512	512 × 512	512 × 512	512 × 512
Stack size (slices)	1	1	1	10
Length (frames)	100	100	100	100
Density (low, medium, high)	100, 500, 1000	60, 400, 700	100, 500, 1000	100, 500, 1000
SNR (levels)	1, 2, 4, 7	1, 2, 4, 7	1, 2, 4, 7	1, 2, 4, 7
Intersection fraction (%)	0.7, 3.4, 6.8	0.8, 4.9, 8.8	0.5, 2.6, 5.4	0.3, 1.2, 2.4

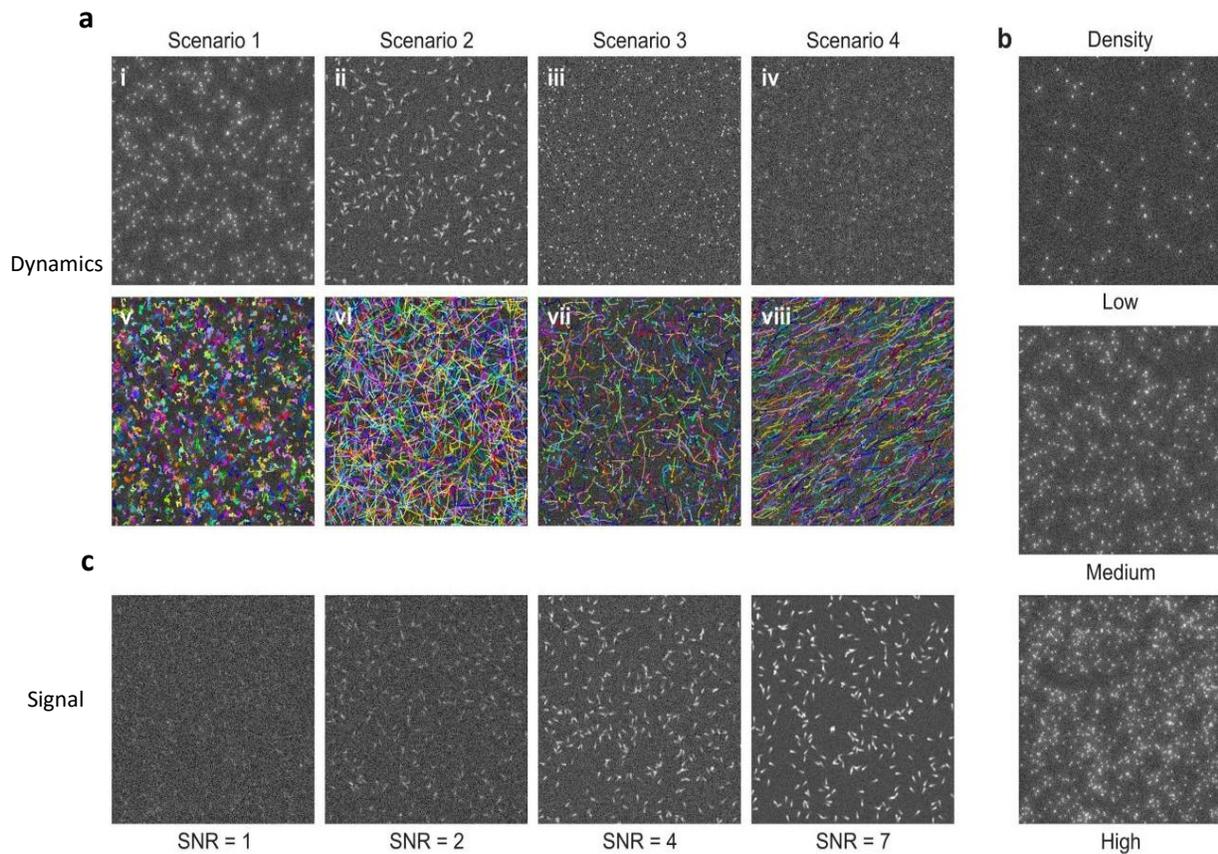


Figure 9: Examples of simulated image data. Representative images of the three main factors (particle dynamics, density, and signal) affecting tracking performance are shown. (a) Four biological scenarios were simulated, of which we show snapshot images (i–iv) and trajectories (v–viii) in arbitrary colors: particles showing random walk motion imaged in 2D+time using widefield microscopy (i, v); larger (elongated) particles represented by asymmetric Gaussians showing directed motion in 2D+time (ii, vi); particles switching between random walk and randomly oriented directed motion imaged in 2D+time using confocal microscopy (iii, vii); and particles switching between random walk and directed motion with restricted orientation imaged in 3D+time (only one slice is shown) using confocal microscopy (iv, viii). (b, c) Three density levels (b; low, medium and high) and four signal-to-noise (SNR) levels (c; 1, 2, 4 and 7) were simulated.

[61]

Table 2
Description of 14 methods used in the comparative study by Chenouard et al. to detect (identify) and link (track) the particles.
[61]

Method	Authors	Detection Method	Linking Method		Dim.	Ref.
			Principle	Remark		
1	Ivo F. Sbalzarini Yuanhao Gong Janick Cardinale	Iterative intensity-weighted centroid calculation	Combinatorial optimization	Greedy hill-climbing optimization with topological constraints	2D & 3D	[62]
2	Craig Carthel Stefano Coraluppi	Adaptive local-maxima selection	Multiple hypothesis tracking	Motion models are user specified (near-constant position and/or velocity)	2D & 3D	[63, 64]
3	Nicolas Chenouard Fabrice de Chaumont Jean-Christophe Olivo-Marin	Maxima after thresholding two-scale wavelet products	Multiple hypothesis tracking	Motion models are user specified (near-constant position and/or velocity)	2D & 3D	[65-67]
4	Mark Winter Andrew R. Cohen	Adaptive Otsu thresholding	Multitemporal association tracking	Post-tracking refinement of detections	2D & 3D	[68, 69]
5	William J. Godinez Karl Rohr	Either thresholding + centroid or maxima + Gaussian fitting	Kalman filtering and probabilistic data association	Interacting multiple models using motion models as specified	2D & 3D	[70, 71]
6	Yannis Kalaidzidis	Lorentzian function fitting to structures above noise level	Dynamic programming	Track assignment based on weighted sum of multiple features	2D	[72]
7	Liang Liang James Duncan Hongying Shen Yingke Xu	Gaussian-mixture model fitting	Multiple hypothesis tracking	Interacting multiple models with forward and backward linking	2D	[73]
8	Klas E. G. Magnusson Joakim Jaldén Helen M. Blau	Watershed based clump splitting and parabola fitting	Viterbi algorithm on state-space representation	Brownian motion is assumed in all cases	2D & 3D	[74, 75]
9	Perrine Paul-Gilloteaux	Either maxima with pixel precision (2D) or thresholding + Gaussian fitting (3D)	Nearest neighbor + global optimization	Global optimization of associations using simulated annealing	2D & 3D	[76, 77]
10	Philippe Roudot Charles Kervrann François Waharte	Histogram based thresholding and Gaussian fitting	Gaussian template matching	Only local and per-trajectory particle linking	2D	[78-80]
11	Ihor Smal Erik Meijering	Gaussian fitting (round particles) or centroid calculation (elongated particles)	Sequential multiframe assignment	Global linking cost minimization	2D	[65, 81, 82]
12	Jean-Yves Tinevez Spencer L. Shorte	Parabolic fitting to localized maxima	Linear assignment problem	Two-step approach (frame-to-frame and segment linking)	2D & 3D	[83, 84]
13	Joost Willemsse Katherine Celler Gilles P. van Wezel	Watershed based clump splitting	Nearest neighbor	Allows merging and splitting of particles and uses a linear motion model	2D & 3D	[85, 86]
14	Han-Wei Dan Yuh- Show Tsai	Morphological opening based clump splitting	Nearest neighbor and Kalman filtering	Essentially a 2D method keeping track of maximum intensity in z	2D & 3D	[87, 88]

The results showed that no one particle tracking method performed best for all data. The best identification methods were based on careful implementation and parameter tuning of any algorithm. The best tracking methods were the ones that used multiframe/multitrack optimization instead of the simpler nearest-neighbour linking. In addition, methods that made explicit use of the prior knowledge about the particle motion in each scenario were more successful than methods that did not. In conclusion, the paper stated the importance of having a “training” data set that simulated real data as accurately as possible. The training data are used to quantitatively evaluate and optimize tracking methods before applying the methods to real data.

One example of a tracking method that uses multiframe/multitrack optimization on Digital Holographic Microscopy (DHM) datasets is the algorithm developed by Sheng et al. [34], who measured the three-dimensional trajectories of two different types of dinoflagellates and their prey. The trajectories were classified by their radius, pitch of helical swimming, translation, and angular velocity. To obtain the trajectories, the author first had to locate the 3D coordinates of the each particle using an automated segmentation method [89]. Then a 3D Lagrangian tracking algorithm was implemented which selected the most likely trajectory based on six criteria. These criteria were the smoothness of the trajectory segment, the smoothness of the 3D velocity, the smoothness and upper bound conditions on acceleration, similarity of cell size, similarity in shape based on correlation between the first image and subsequent images, and the similarity of the aspect ratio between the minimum and maximum dimensions of the in-focus images. More formally, this procedure is known as a multi-layer decision tree combined with a support vector machine classifier. The most likely trajectories were examined manually and the

most correct-looking one selected. Even though this method can result in accurate tracks, it is very time consuming and only appropriate for tracking a small number of microorganisms within a sample.

In this thesis, a high precision machine-learning particle identification/detection algorithm based on linear logistic regression [90] was implemented. The algorithm requests an expert user to identify bacteria from a training data set, which is a small subset of the recorded dataset. Therefore, the training data is as close as possible to the rest of the recorded dataset. This training step would account for the large variability in reconstruction image settings among different DHM data sets.

As for particle tracking/linking, the simple nearest-neighbour Hungarian linking algorithm developed by Tinevez et al. [91] was implemented here. Links are created amongst particle pairs found to be the closest in Euclidean distance. By virtue of the Hungarian algorithm, it is ensured that the sum of the pair distances is minimized over all particles between two frames. The main advantage of using this simple linking algorithm is that it does not require any prior knowledge about the microorganism's movements throughout the data set. Since the particle identification/detection step described previously has high precision, accurate trajectories were obtained using one of the simplest tracking/linking algorithms available.

In the chapter 4 of this thesis, the performance of the algorithm is evaluated using two "gold standard" DHM datasets in which the microorganisms were manually labelled and tracked by an expert.

CHAPTER 3: METHODS AND EXPERIMENTAL TECHNIQUES

3.1 Data Acquisition

SPECIFICATIONS AND DESIGN OF THE DIGITAL HOLOGRAPHIC MICROSCOPE

The Digital Holographic Microscope (DHM) used in this study was designed and built by Wallace et al [92]. It is a twin-beam off-axis DHM, suitable for extreme environments in terms of mechanical and thermal stress. Figure 10 shows a schematic of this DHM in its laboratory implementation. Figure 11 shows the same microscope, but in its robust, field portable implementation. Table 3 lists the properties and technical specifications of this DHM.

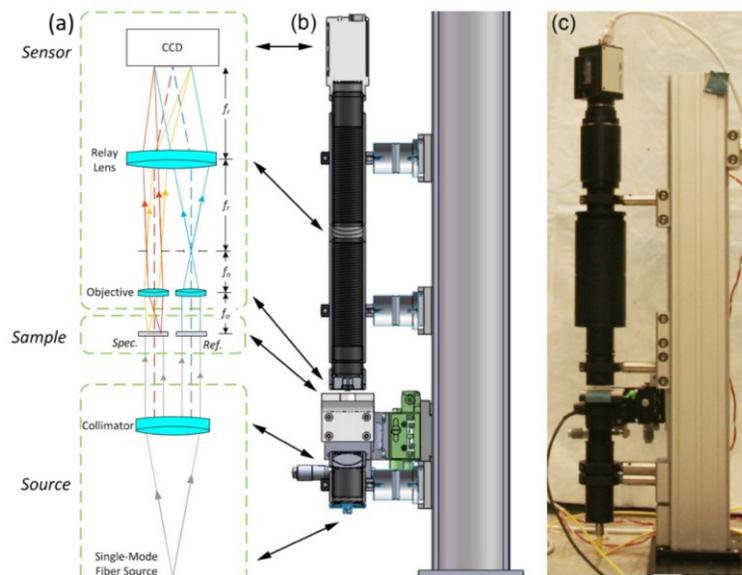


Figure 10

Reprinted with permission from ref [92], Optics Express:

Schematic and images of the compact, twin-beam digital holographic microscope in its laboratory implementation.

- (a) Schematic showing four main elements (discussed in the text): the source, the sample (specimen path is labeled Spec. and reference path is labeled Ref.), the microscope, and the sensor.
- (b) Solid model of the hardware. The fiber-fed source assembly is at the bottom, and the imaging camera is at the top. The microscope optics – comprised of the two aspheric lenses and the relay lens – are contained within the 300 mm long lens tube. The three-axis stage between the source the microscope optics provides easy manual manipulation of the specimen under study.
- (c) Photograph of the instrument in the laboratory. A shutter over the collimating lens protects against condensation, and is operated remotely by a controller.

[92]



Figure 11
Field implementation of the portable DHM designed by Wallace et al. [93]

Table 3
Reprinted with permission from ref [92], Optics Express:
Properties of the DHM designed by Wallace et al. [92]

Property	Value	Unit	Note
Operating Wavelength	405	nm	Single-mode fiber-coupled laser
Objective focal length f_0	7.6	mm	Aspheric singlet
Objective Numerical Aperture	0.30		
Relay lens focal length f_r	150	mm	Achromatic Doublet
System magnification	19.7		
Lateral resolution	0.7	μm	
CCD pixel size	3.45 x 3.45	$\mu\text{m} \times \mu\text{m}$	2448 x 2050 CCD chip
Sample imaging volume	360 x 360 x >600	$\mu\text{m} \times \mu\text{m} \times \mu\text{m}$	In 2048 x 2048 (4Mpx) mode
Sampling Rate	15	Frames per second	4Mpx mode; 22 fps with 1Mpx
Instrument length	400	mm	Input fiber to back of CCD

SAMPLE CHAMBER

A microfluidic sample chamber was designed to fit the DHM without leaking. The sample chamber contained two channels, each channel holding a volume of 24 μ L. One channel was for the sample and the other channel was for a fluid-only reference, as required by DHM. The chambers allowed the laser light to pass through high-quality glass as shown in Figure 12 . The sample chamber is disassembled and cleaned before every use.

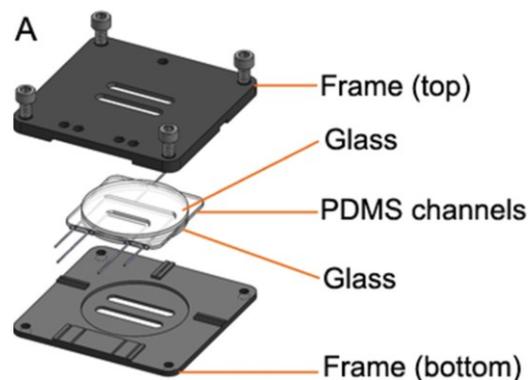


Figure 12
Sample chamber design schematic used in this thesis.
PDMS: polydimethylsiloxane, a common material used for construction of microfluidic channels.
[94]

PREPARATION OF MICROBIAL SAMPLES

The *Bacillus subtilis* cells were grown to mid-log phase in Lysogeny Broth (LB) in a shaking incubator at 30 °C. They were then diluted in motility medium (10mM potassium phosphate, 10mM NaCl, 0.1mM EDTA, 0.1mM glucose, pH 7.0) immediately before being inserted into the sample chamber and imaged using the DHM at room temperature.

The *Colwellia psychrerythraea* cells were maintained in ½ strength 2216 Marine Broth (Difco) at 6°C. They were then diluted using the same Difco broth immediately before being inserted into the sample chamber and imaged using the DHM at room temperature.

RECONSTRUCTION OF A HOLOGRAM INTO MULTIPLE 2D Z-PLANES

The commercial KOALA® software from LynceeTec was used for the holographic reconstructions [95]. For each hologram imaged at a specific time-frame, two series of images were reconstructed that spanned multiple z-planes; the intensity images and the phase images. The images were saved in 8-bit .tiff format; each image had a unique name which contained both its z slice and time-frame location. This process was repeated for each hologram imaged at all time-frames. The holograms of “gold standard 1” and “gold standard 2” datasets were numerically reconstructed at a z spacing of 1.25 µm/slice and 2.50 µm/slice, respectively.

3.2 Generating Motility tracks

OVERVIEW

This chapter presents an Object-Identification Learning Algorithm suitable for a large variety of 4-D datasets. The user identifies objects of interest within a subset of the dataset, and then similar objects within the entire dataset are presented back to the user by the algorithm.

Generating motility tracks is achieved in two steps: the first step is Object Identification and the second step is Object Tracking.

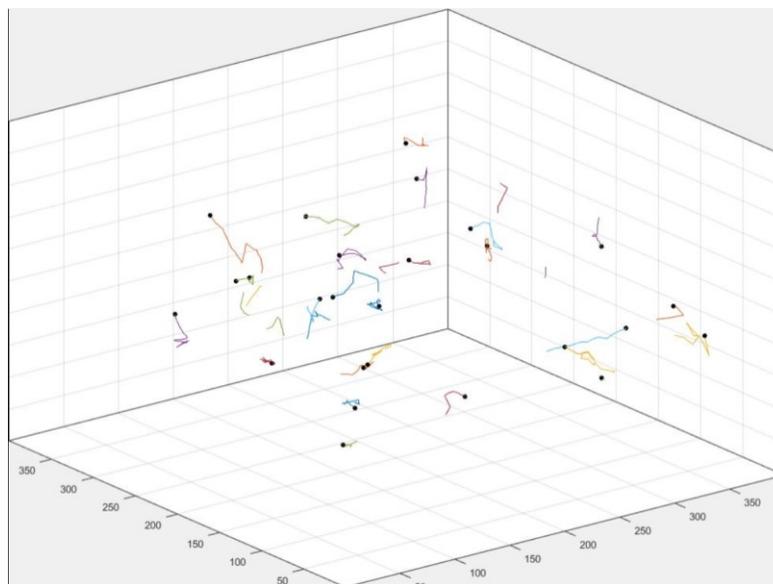


Figure 13
3D tracks of *Vibrio alginolyticus* moving through a sample volume of size $360 \times 360 \times 200 \mu\text{m}^3$.
These tracks were obtained using the 3D Object-Identification Learning Algorithm described in this paper.

Table 4
Generalised input and output of the algorithm described in this study.

Input	4D holographic time series reconstruction dataset.
Output	3D Tracks of objects within the dataset.

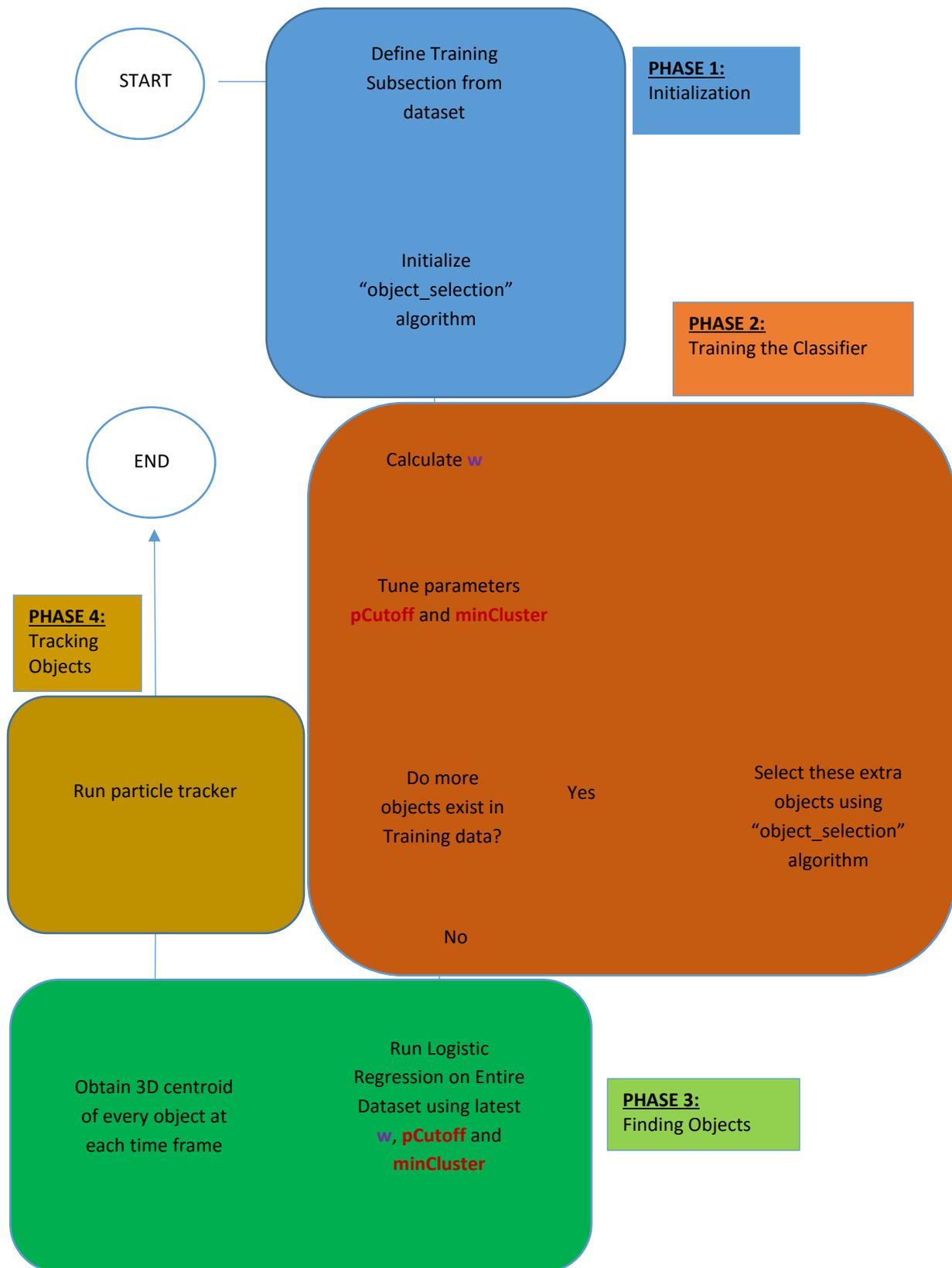


Figure 14
Algorithm Flowchart

3.2.1 - OBJECT IDENTIFICATION

Goal: To find the 3D centroid of every object at all time frames within any 4D holography dataset.

This section will describe a binary classification algorithm used to separate pixels representing microorganisms from pixels representing background within a 4D holography dataset. This Identification algorithm outputs the 3D centroids of every microorganism in the dataset D.

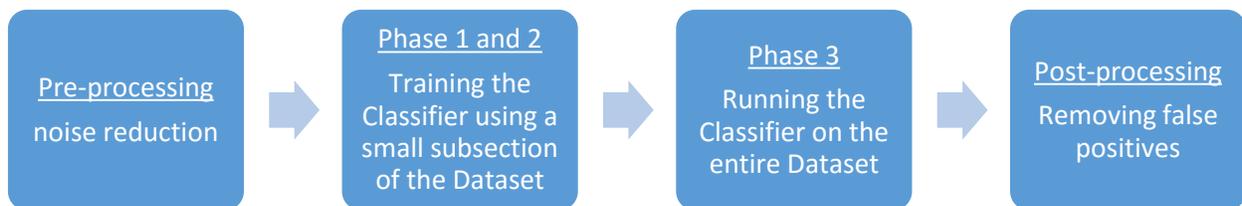


Figure 15
Object Identification Methodology Overview

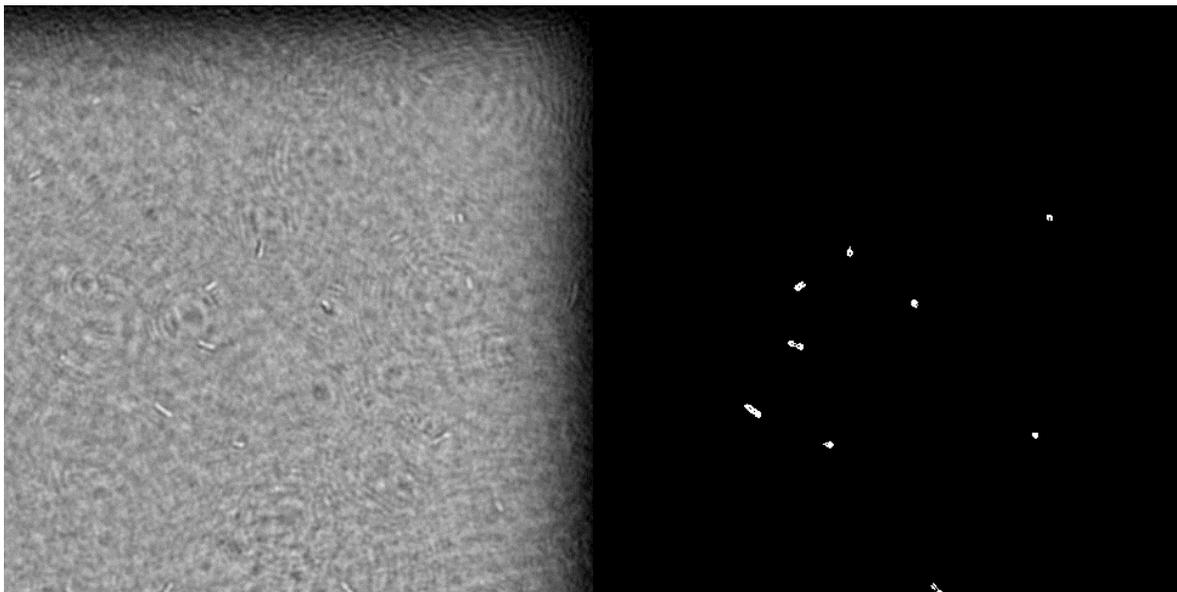


Figure 16
Left Image: raw hologram phase reconstruction image.
Right Image: Pixels classified as bacteria (white pixels) and pixels classified as background (black pixels) of the Left Image.

Variables and parameters

Table 5
Variable and Parameter list of the object identification algorithm

Input Variables	Variable Type	Description
D	4D 8-bit positive integer Matrix	Variable D is a 4D hyper-stack of two dimensional grayscale images that pan through the sample depth and time.
D_{train}	3D 8-bit positive integer Matrix	D_{train} is a 3D stack of two dimensional images at 1 point in time. This stack represents the entire volume at 1 time-frame.
Parameters	Parameter Type	Description
p_{Cutoff}	Double-precision number between zero and one	p_{Cutoff} is the value representing the threshold minimum probability required for a pixel to be classified as “object”.
minCluster	Positive 8-bit integer	minCluster is the minimum number of connected pixels that an “object” could possess. Connected “object” pixels of a size below minCluster are ignored and treated as noise.
Output Variables	Variable Type	Description
Points	Double-precision Cell Array	Output Variable Points is a cell array containing all the 3D object centroids found within every time-frame.

Linear Logistic Regression Classifier

A linear logistic regression classifier [90] can be used for any dataset with a binary output. It is trained using a supervised training dataset; supervised meaning that an answer key exists for the training dataset.

The Linear Logistic Regression pixel classifier described in this section is flexible; therefore, it can deal with the large variability of DHM datasets. It is implemented in MATLAB using the `glmfit()` and `glmval()` functions found in MATLAB's "Statistics and Machine Learning" toolbox. It is a supervised learning algorithm.

The classifier $h_w(x)$ is calculated to map the pixel features matrix X_{train} (see Figure 14) of the pixels in the training data set D_{train} , to a binary vector y_{train} . w is the linear weights of the feature columns in X_{train} .

$$D_{train} \rightarrow X_{train}$$

$$X_{train} * h_w(x) = y_{train}$$

$$h_w(x) = \frac{1}{1 + e^{w^T x}} \quad (3.1)$$

X_{train} and y_{train} are known, and therefore the only unknown is the linear weights vector w of the classifier $h_w(x)$. w cannot be found in closed form, hence a gradient descent approach is necessary to calculate w . α is the gradient-descent learning-rate parameter.

$$w \leftarrow w + \alpha \sum_{i=1}^m (y_{train,i} - h_w(x_i)) x_i$$

$$w \leftarrow w + \alpha [X_{train}]^T (y_{train} - \hat{y}) \quad (3.2)$$

Once w is calculated, the classifier $h_w(x)$ is then applied to the feature vector X of the entire dataset D . This operation equals a vector y , corresponding to the probability of a pixel being bacteria.

$$D \rightarrow X$$

$$X * h_w(x) = y$$

$$0 < y < 1 \tag{3.3}$$

The threshold minimum probability required for a pixel to be classified as “object” is the parameter p_{cutoff} . The threshold minimum size of 3D connected pixels required for a cluster of “object” pixels to be classified as microbe is the parameter $minCluster$.

Clusters of pixels that satisfy these two criteria are identified as cells, and their location coordinates are recorded.

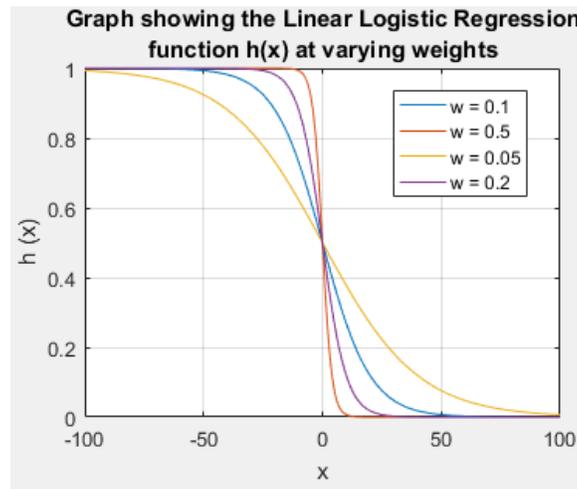


Figure 17

Linear Logistic Regression Classifier function. The output varies smoothly between 0 and 1, making this function ideal for calculating probabilities that also lie between 0 and 1.

Three dimensional pixel Features

Table 6 describes the 3D pixel features used to classify the dataset. The functions used to extract these features can be implemented in the GPU using MATLAB's built-in toolboxes.

Table 6
Pixel Features used in the classifier

Pixel Feature	Relevance
Absolute difference between Pixel Grayscale Value and median pixel value (the median pixel grayscale value is set to 127 in preprocessing).	Objects are usually brighter or darker than the median grayscale value.
Median of pixel neighbourhood	Taking the median will account for noisy outliers within the pixel region
Standard deviation of pixel neighbourhood	Noisy pixel neighbourhoods will have a high standard deviation
Gradient of the pixel	The gradient is high at the boundary between an object and the background
Absolute difference between grayscale value of the current pixel and the pixel above it.	In phase reconstructions, the objects' grayscale values fluctuate heavily between very bright and very dark around the z-slice where the object is located.
Absolute difference between grayscale value of the current pixel and the pixel below it.	In phase reconstructions, the objects' grayscale values fluctuate heavily between very bright and very dark around the z-slice where the object is located.
Standard Deviation of the entire current 2D z-slice image	The standard deviation of an image is low when some objects are in focus, and high when objects are not in focus.

Object Identification Methodology

Pre-processing

Once holograms are acquired and reconstructed onto separate z-planes, they undergo four steps before being input into the Identification Algorithm.

The first step is to shift the grayscale values of every reconstructed image such that the median pixel grayscale value is 127. This is to ensure that overexposed and underexposed images become uniform in brightness and therefore, numerically comparable with each other.

The second step is to subtract the median background from every reconstructed image within the dataset. A window containing the current time frame, the temporal previous 5 frames and the temporal next 5 frames is created for every time step. The background is defined as the median grayscale value of each pixel location within each window. These background images are then subtracted from dataset.

Running the background subtraction algorithm is time consuming. It would be faster and easier to generate only one background for every z-slice, rather than the method described above which generates a unique background for every time frame and every z-slice. This trade-off is ultimately worth it because the background subtraction algorithm produces the most accurate results by also removing artifacts within the dataset that drift very slowly over a long period of time, such as dead cells.

The third and final step is to increase the contrast of the image, making darks darker and lights lighter. Increasing the contrast makes training the classifier easier for the user, because the bacteria become more visually apparent. This is achieved by calculating the difference between

the median grayscale value (127) and the image's pixel values, and multiplying that difference by a positive factor cf , where $cf > 1$. Depending on the dataset, cf usually varies between 2 and 5.

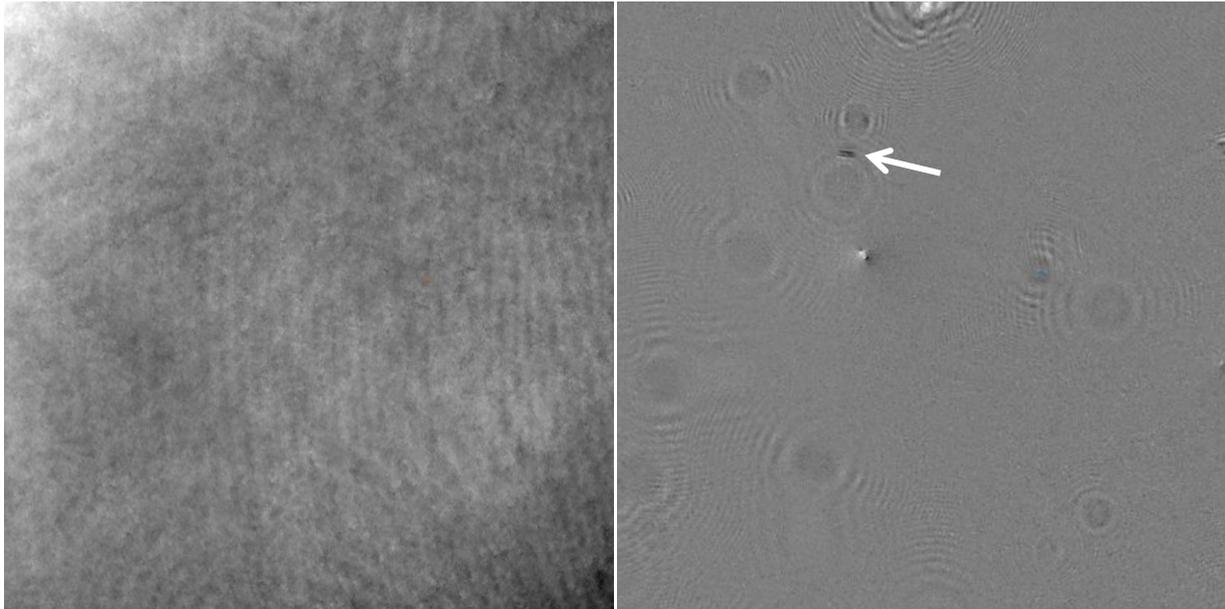


Figure 18

Left Image: Phase reconstruction image of a hologram at a specific z slice and time frame
Right Image: The same image as the left image after pre-processing. The median background has been removed from the image and the contrast has been doubled by setting $cf = 2$. Patterns of concentric circles represent moving objects within the dataset that are out of focus; one dark rod-shaped cell is within this z plane of focus (arrow).

Phase 1 – Initialization

In this phase of the algorithm, the user must define a training dataset D_{train} , and initialize the selection algorithm.

The training dataset is a subset of the entire holography dataset. It must contain instances of positive objects. A good training dataset could be the entire volume of the sample at a single time frame. For shallow samples, multiple time frames may be used if desired.

To initialise the selection algorithm described in Phase 2, the user must scan through the entire training dataset, and record the (x,y,z,t) location of a pixel within each cell.

A binary answer key vector $y_{train,0}$ is initialized; user labelled bacteria pixels are assigned the value 1, and all other pixels are assigned the value 0.

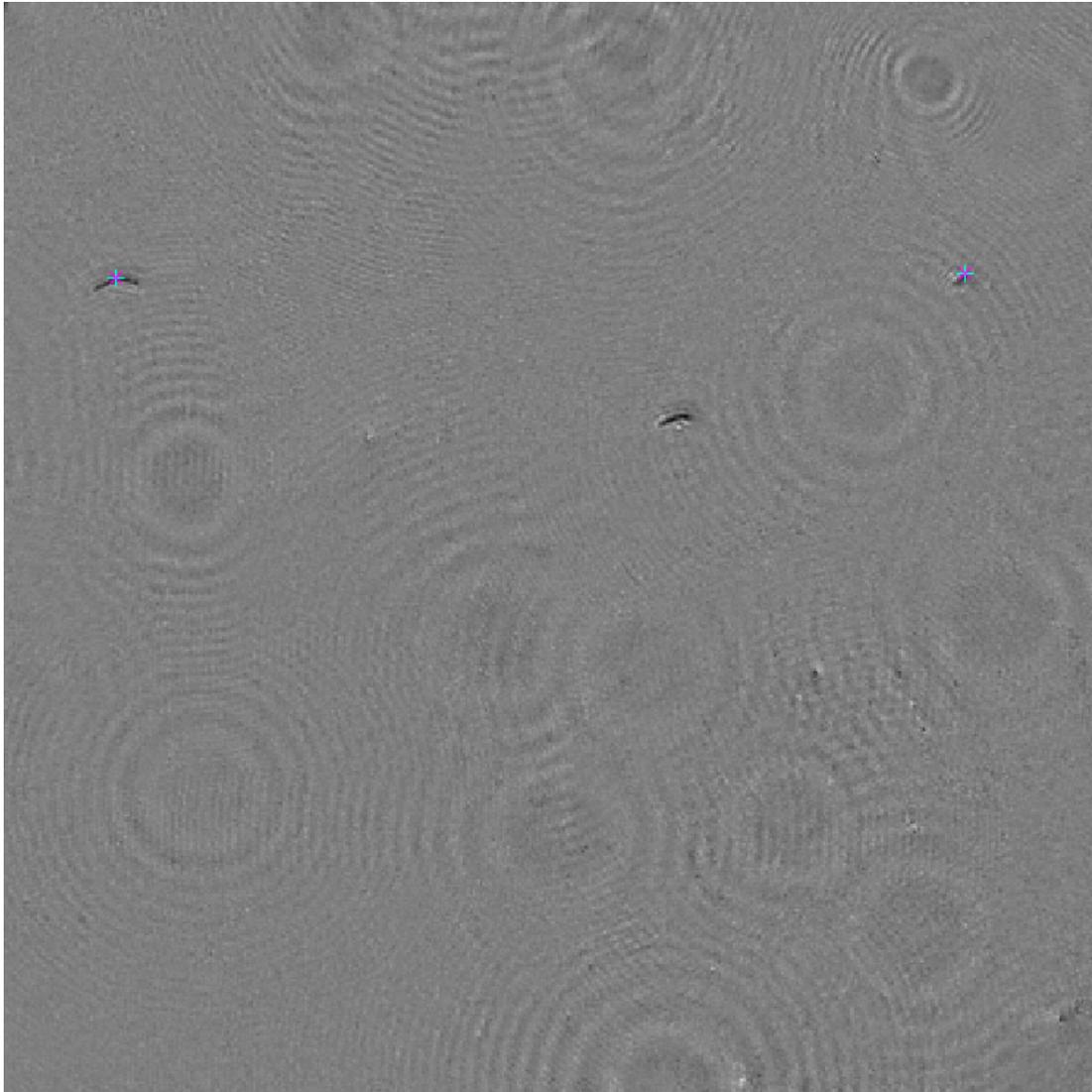


Figure 19 – Phase 1

User-selected cell's pixels (blue-purple stars) at one z-slice within the training dataset. These coordinates will be used to initialize the selection algorithm described in Phase 2, which will make selecting more pixels easier and faster than manual labelling.

Phase 2 – Training the Classifier

Training the logistic regression classifier is achieved in two steps.

The first step is to complete the binary answer key vector y_{train} , described at the end of Phase

1. The second step is to calculate the weight vector w for the classifier $h_w(x)$.

Vector y_{train} is updated iteratively. Each iteration labels a new group of pixels as bacteria, using the selection algorithm described in this section, until all bacteria pixels within the training dataset have been labelled.

The first iteration of the weight vector w_1 is calculated using X_{train} and the initial $y_{train,0}$ vector created at the end of Phase 1. Then D_{train} is classified using $h_{w_1}(x)$. Parameters p_{cutoff} and $minCluster$ are tuned (see Table 5), and the resulting classified image is presented.

p_{cutoff} is the value representing the threshold minimum probability required for a pixel to be classified as belonging to a cell. Its value is between zero and one. $minCluster$ is the minimum number of connected pixels that an “object” could possess. Connected “object” pixels of a size below $minCluster$ are ignored and treated as noise.

From this resulting classified image, a cell’s pixels are identified by the user. These pixels are labelled as belonging to a cell and hence their value is set to 1 in vector $y_{train,1}$.

This process is repeated until all bacteria pixels within the training dataset have been labelled and thus, the vector y_{train} is complete.

Once all the bacteria pixels have been selected and recorded in vector y_{train} , the final value of weight vector w is calculated, followed by an appropriate selection of parameters p_{cutoff} and $minCluster$.

Training is now complete; the variables and parameters found at the end of phase 2 are passed to phase 3.

$$w_n \leftarrow w_n + \alpha [X_{train}]^T (y_{train,n-1} - h_{w_{n-1}}(x_i))$$

$$X_{train} * h_{w_n}(x) = q_n \quad (3.4)$$

for \forall elements i in vector $q_n(i)$:

if $q_n(i) > p_{cutoff}$,

$y_{train,n}(i) = 1$;

else

$y_{train,n}(i) = 0$;

end

end

(3.5)

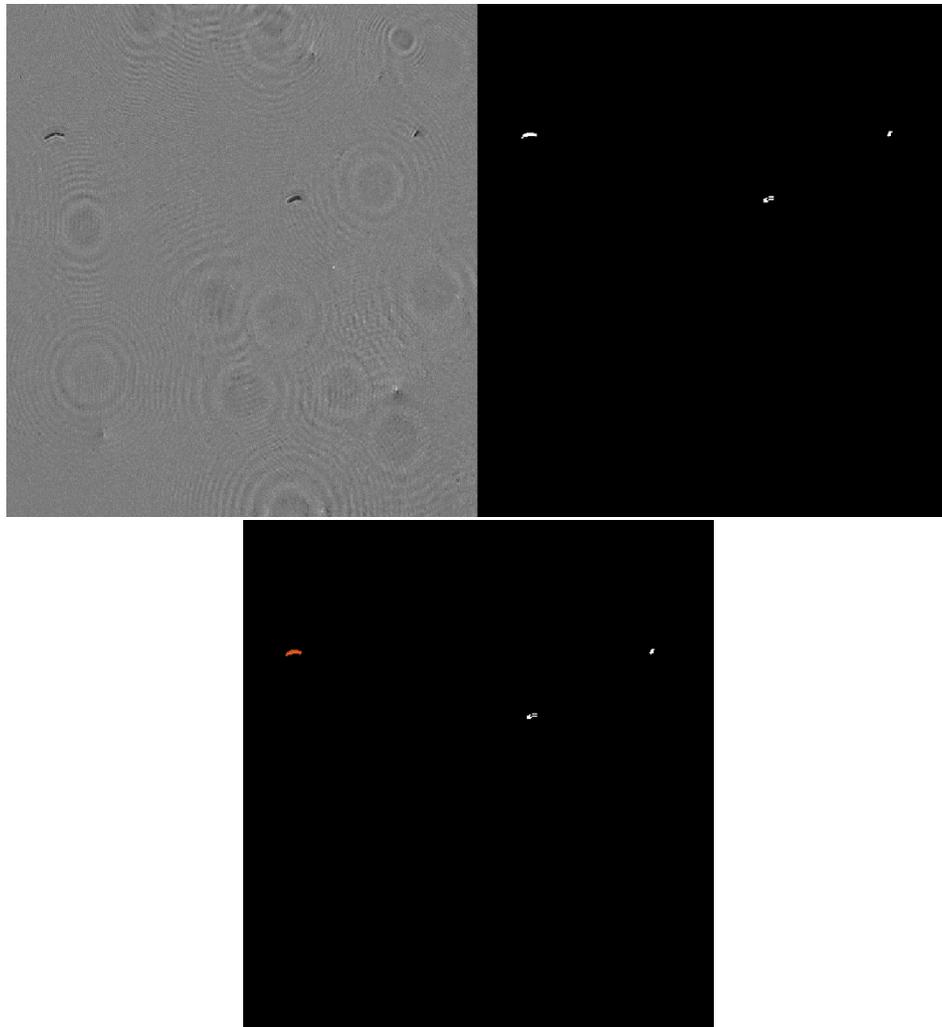


Figure 20 – Labelling a new group of pixels as bacteria using the training classifier.

Top Left: original phase reconstruction image.

Top Right: segmented image with white pixels labelling bacteria, and black pixels labelling background.

Bottom: User selected an extra cluster of unlabelled connected bacteria pixels, to be added to the binary answer key vector y_{train} .

This process is repeated until all pixels of bacteria are labelled as such.

Phase 3 – Finding objects within the dataset

The linear logistic regression algorithm described previously is implemented on all pixels of the entire dataset. The linear weights vector w of the classifier $h_w(x)$ was calculated from the training dataset at the end of Phase 2. Parameters p_{cutoff} and $minCluster$ were also defined at the end of Phase 2.

All the 3D connected pixel objects are found at every time frame. The threshold for minimum number of 3D connected pixels in an object to be considered a microbe is defined as *minCluster*.

The 3D Centroids of objects considered microbes at each time frame are recorded and stored in output variable *points*.

Post-processing

The purpose of this step is to remove false positives.

A characteristic of false positive points is that they appear for a short duration of time and then disappear. Points with this characteristic are removed using the following method:

1. Define a minimum track duration threshold.
2. Run a simple point tracker algorithm based on nearest neighbour identification, called 'simple_tracker.m' [91].
3. Remove points that belong to tracks for a duration below the threshold.

3.2.2 - GENERATING TRACKS

Goal: To link 3D centroids together from consecutive time frames to generate a Track. Each track describes the trajectory of a microbe throughout the sample volume.

Object tracking/linking is the 4th and final Phase of the algorithm described in this thesis. This section will describe the linking algorithm used to plot the trajectories of the centroids obtained from the Identification algorithm described in section 3.2.1

Variables and Parameters

Table 7
Variables and Parameters list of the object tracking algorithm

Input Variables	Variable Type	Description
Points	Double-precision Cell Array	Variable Points is a cell array containing all the 3D object centroids found within every time-frame.
Parameters	Parameter Type	Description
max_linking_distance	Positive number	Defines a maximal distance for particle linking. Two particles will not be linked (even if they are the remaining closest pair) if their distance is larger than this value.
max_gap_closing	positive 8-bit integer	Defines a maximal frame distance in gap-closing. Frames further way than this value will not be investigated for gap closing.
Output Variables	Variable Type	Description
Tracks	Unsigned 8-bit integer cell array	One cell in cell array Tracks per found object track. The integer array in each cell shows the index of each object in Points within the track.

Object Tracking Methodology

Phase 4 – Tracking the objects' movements

Tracking the centroids found after implementing the Object Identification Algorithm presented earlier is achieved by running a simple tracker based on nearest neighbour identification, called 'simple_tracker.m'[91].

Appropriate values for parameters *max_linking_distance* and *max_gap_closing* must be chosen. *max_linking_distance* is the maximum Euclidean distance for two points at consecutive time frames to be considered the same point travelling through a trajectory. *max_gap_closing* is the maximum number of dropped frames within a single track.

3.2.3 - VALIDATION

Validation Metrics

The $F_{0.5}$ score was calculated to validate the performance of the algorithm against a gold standard validation dataset (see section 'Gold Standard' below).

The $F_{0.5}$ score is defined as:

$$F_{0.5} = (1 + 0.5^2) \cdot \frac{P \cdot R}{(0.5^2 \cdot P) + R} \quad (3.6)$$

Where precision (P) and recall (R) are defined as:

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN} \quad (3.7)$$

Where (TP) is true positives, (FP) is false positives and (FN) is false negatives.

The $F_{0.5}$ score was chosen for validation to weigh precision higher than recall, by attenuating the influence of false negatives. The rationale behind this is that this algorithm was developed to precisely find at least 50% of the microbes within the dataset, and not to find all the microbes within the dataset. Therefore, False negatives are acceptable more so than false positives.

The output centroids found using the Identification algorithm were validated against the Gold Standard centroids described in the section below. An identification tolerance was defined as the maximum Euclidean distance between a centroid found using the identification algorithm, and the actual location of that centroid found in the Gold Standard.

The tracks returned by the Tracking algorithm were validated against the Gold Standard tracks described in the section below. A tracking tolerance was defined as the minimum percentage of points found within a track to be ‘detected’ by the tracking algorithm.

Gold Standard

Two DHM validation datasets were created with the properties listed in Table 8. These properties were measured manually using an expert human. They were defined as the Gold Standard towards which the Identification Algorithm described here would be validated. Figure 21 and Figure 22 show three-dimensional plots of the tracks within each of the two Gold Standard validation datasets.

Table 8
Properties of the Gold Standard Datasets used for validation

Property	Validation Dataset 1	Validation Dataset 2
Sample volume	360 x 360 x 252 μm^3	360 x 360 x 392 μm^3
Object	<i>Bacillus subtilis</i>	<i>Colwellia psychrerythraea</i>
Concentration	≤ 10 cells per sample volume	≤ 10 cells per sample volume
Object size, shape	$\sim 8 \mu\text{m}^3$, elongated	$\sim 2 \mu\text{m}^3$, comma-shaped
Resolution of each 2D Hologram Reconstruction Images'	454 x 452 pixels	512 x 512 pixels
Number of z-plane Reconstructions	201	157
Number of unique time frames	84	18
z-plane resolution	1.25 $\mu\text{m}/\text{slice}$	2.50 $\mu\text{m}/\text{slice}$
Total number of objects in the dataset	324	98
Total number of unique tracks in the dataset	8	6

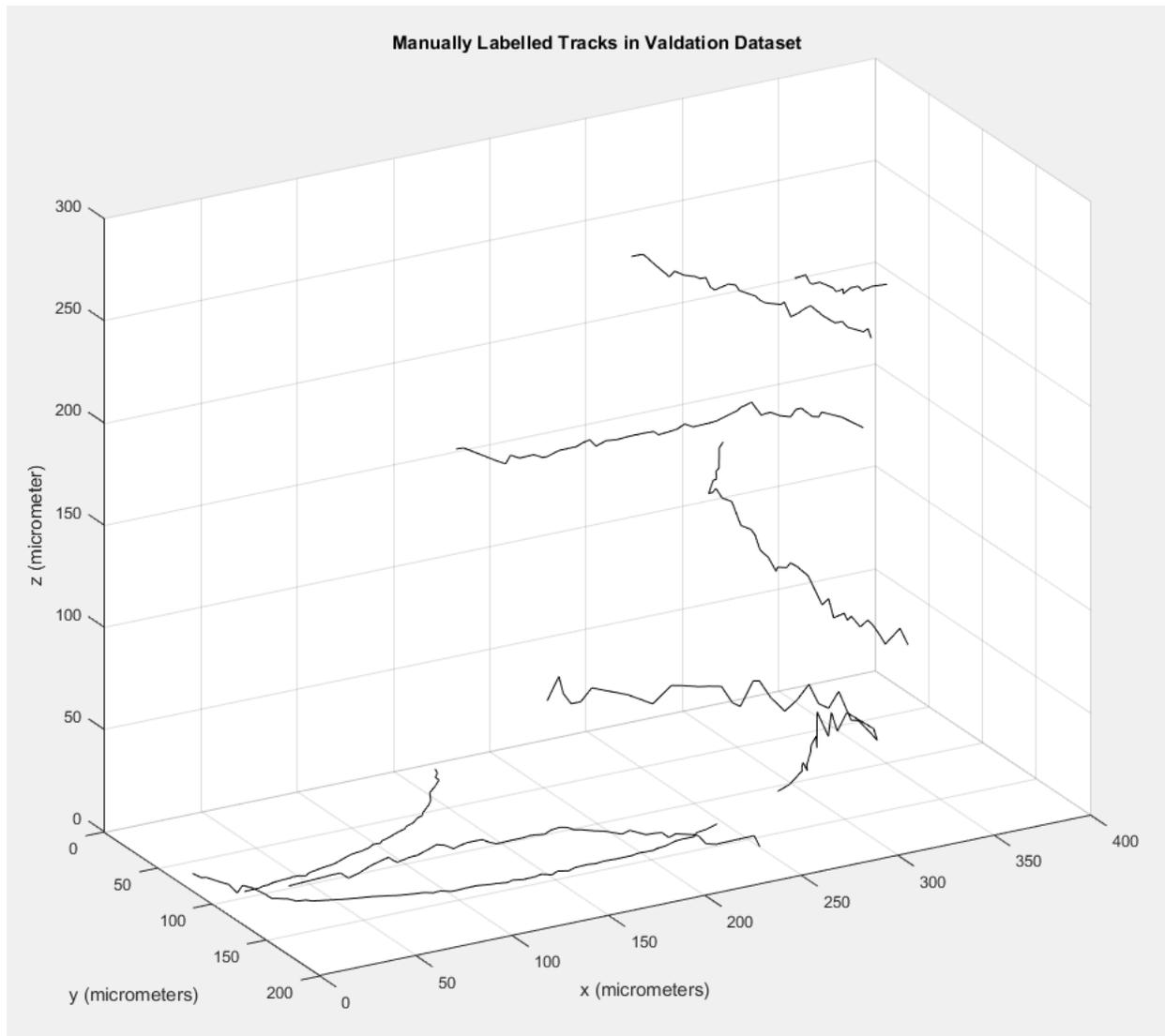


Figure 21

The 8 Microbial *B. subtilis* Tracks (solid black lines) within Validation Dataset 1, found manually by the expert human user.

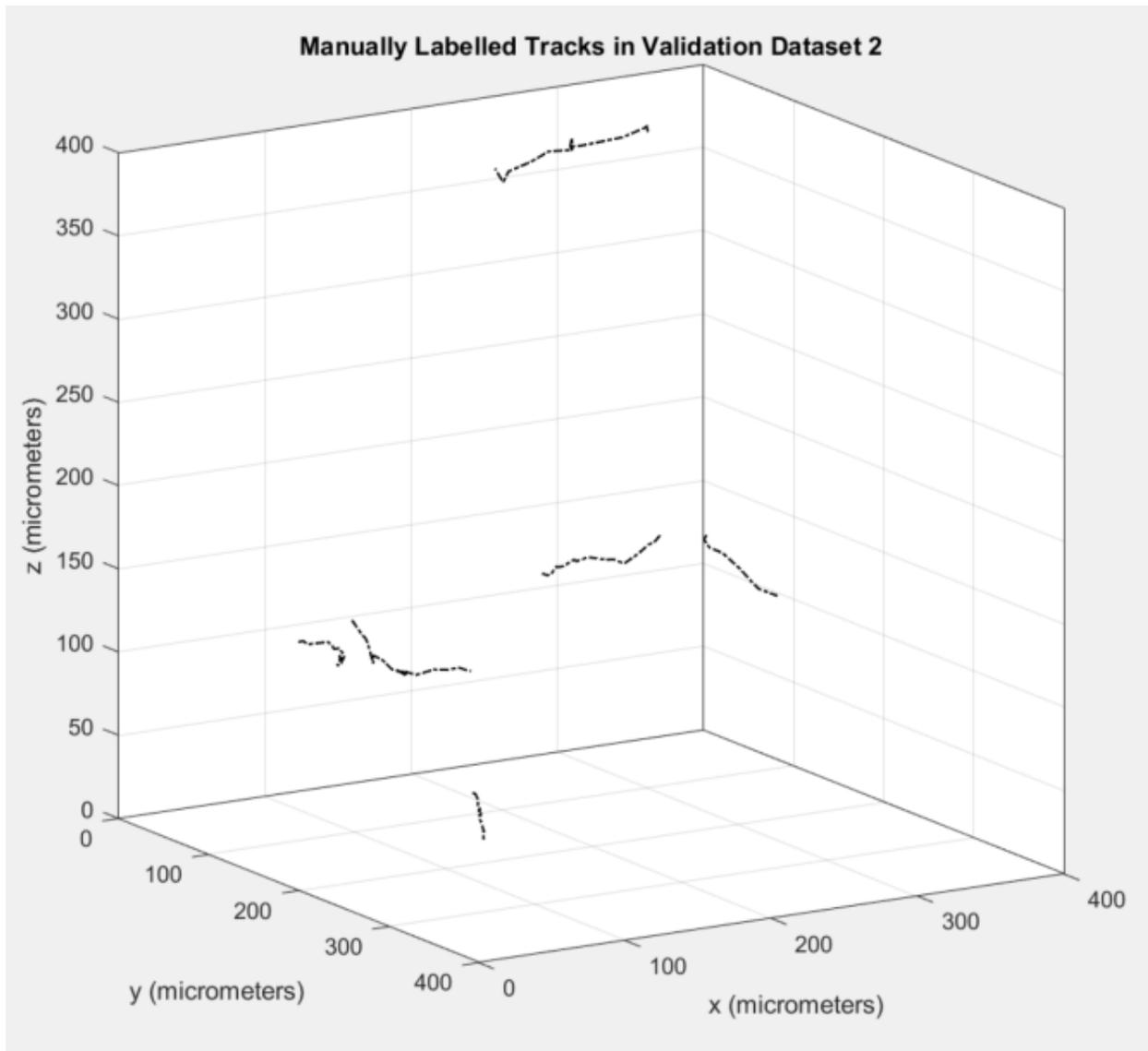


Figure 22

The 6 *Colwellia psychrerythraea* tracks (dashed black lines) within Validation Dataset 2, found manually by the expert human user.

3.3 Optimizing Computational Speed

OPTIMIZING RAM USAGE

Running this algorithm presents challenges to the user because of the large 4D dataset. In order to reduce the memory usage, all the images are resized from 2048 x 2048 pixels to 512 x 512 pixels. Only a single time-step's 3D volume is loaded into memory at any given time. Only when the 3D coordinates of the object centroids found using this algorithm are saved in the output coordinates matrix, then the 3D volume of the next time step is loaded into the same memory location as the previous 3D volume.

USING A GPU FOR IMAGE PIXEL OPERATIONS

Most of the features for logistic regression used in this algorithm are standard image pixel operations, such that they could easily be implemented in the Graphical Processing Unit (GPU) using standard image processing routines. This effectively reduces the computational time required to completion.

CHAPTER 4: RESULTS AND OBSERVATIONS

Overview

Two validation datasets of different properties were used to evaluate the performance of the Particle Identification and Tracking algorithm described in section 3.3. Validation dataset 1 contained eight *B. subtilis* microorganisms of various orientations and sizes, and validation dataset 2 contained six *Colwellia psychrerythraea* microorganisms of various orientations and sizes.

The *B. subtilis* microorganisms in validation dataset 1 were elongated and rod shaped, from 5 μm to 12 μm in length. The *Colwellia psychrerythraea* microorganisms in validation dataset 2 were also rod-shaped but less elongated than *B. subtilis*. They were smaller, ranging in length from 1 μm to 3 μm . Some of these microorganisms are shown in Figure 23 and Figure 24 .

4.1 Particle Identification Algorithm Results

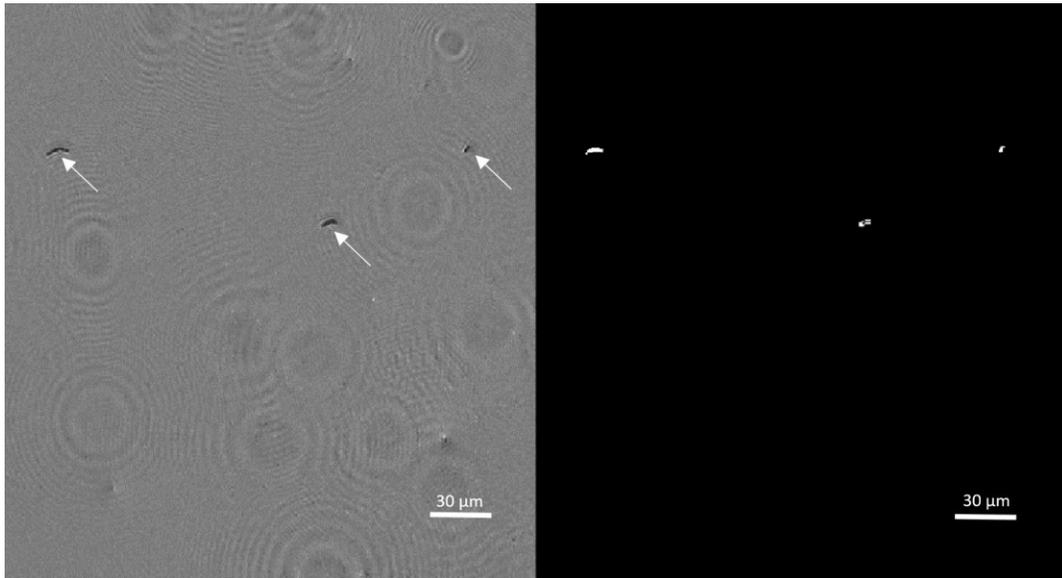


Figure 23

Left Image: one z-slice containing three *B. subtilis* microorganisms (white arrows) within the 3D training volume of Validation Dataset 1.

Right Image: Post-Classification Binary Image of the Left Image. White pixels represent microorganisms within the plane of focus, and black pixels represent background.

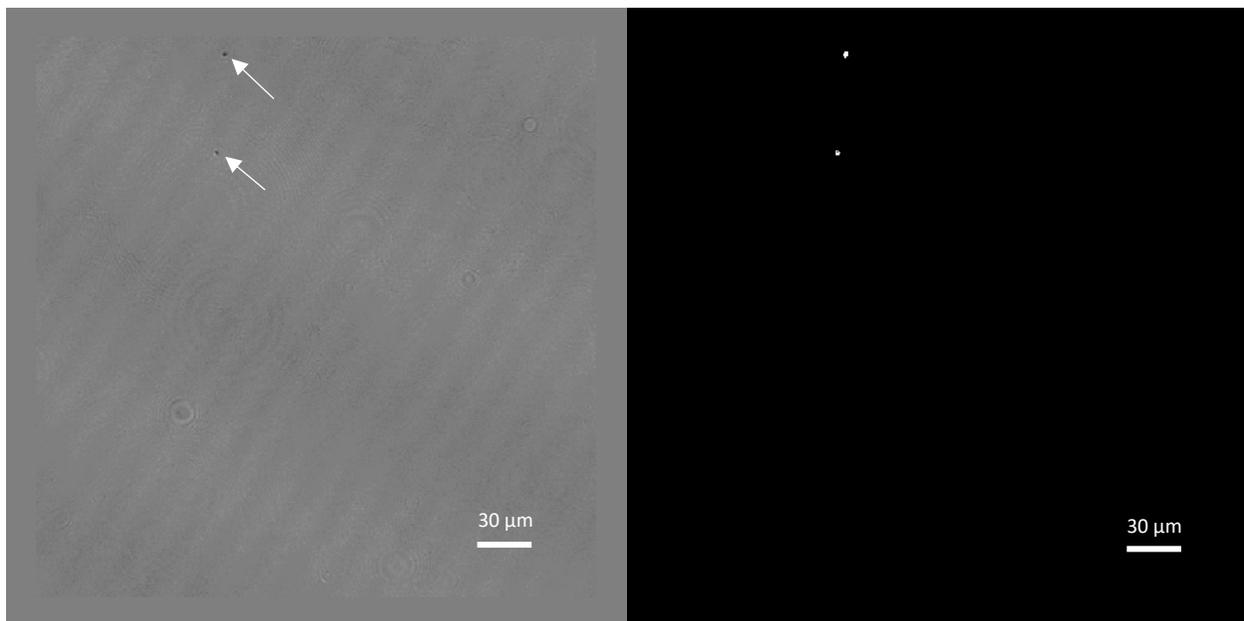


Figure 24

Left Image: one z-slice containing two *Colwellia psychrerythraea* microorganisms (white arrows) within the 3D training volume of Validation Dataset 2.

Right Image: Post-Classification Binary Image of the Left Image. White pixels represent microorganisms within the plane of focus, and black pixels represent background.

PARTICLE IDENTIFICATION RESULTS FROM VALIDATION DATASET 1

Overview

The algorithm was trained using the validation dataset 1 volume at time-frame $t = 85$. This volume contained 4 *B. subtilis* microorganisms. Figure 23 displays a single z-slice of this training volume, containing 3 of the 4 microorganisms in the training dataset.

The weight vector w was calculated using the feature matrix X_{train} and the user-labelled binary identification vector y_{train} . Parameters $pCutoff$ and $minCluster$ were tuned to identify all groups of microorganism pixels within this sample volume.

Table 9
Parameter values assigned using Training dataset of Validation Dataset 1.

Parameter name	Assigned value
$pCutoff$	0.07
$minCluster$	60

The calculated weight vector w was used to classify all the pixels in the entire validation dataset1, and find the microorganisms present at all time frames. The results of the microorganism identification algorithm are presented in Figure 25 as a three-dimensional plot, and in Figure 26 as a top-view two-dimensional plot.

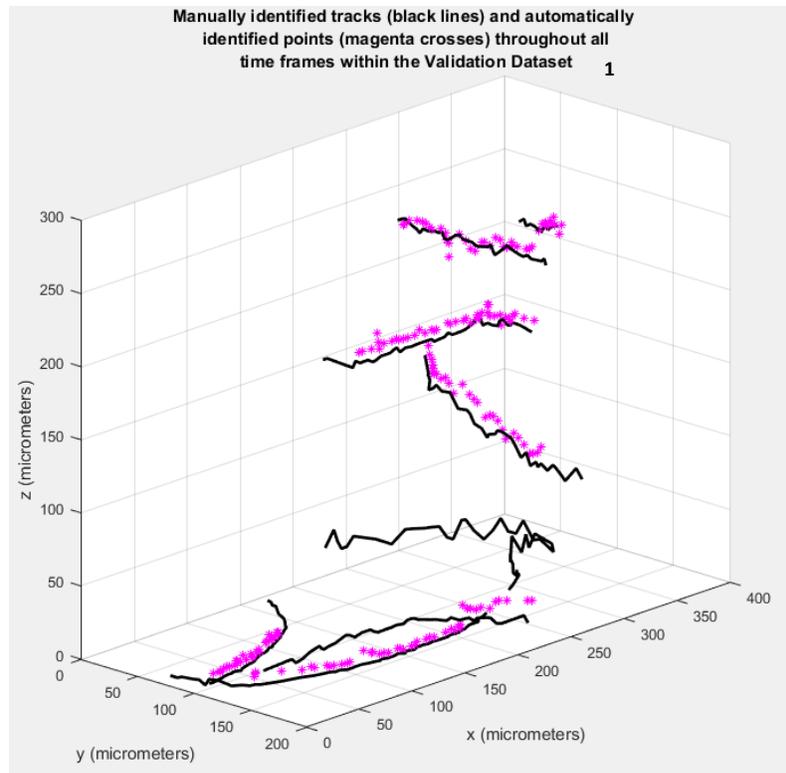


Figure 25
3D plot of manually identified tracks (black) and automatically identified points (magenta) throughout all time frames within Validation Dataset 1.

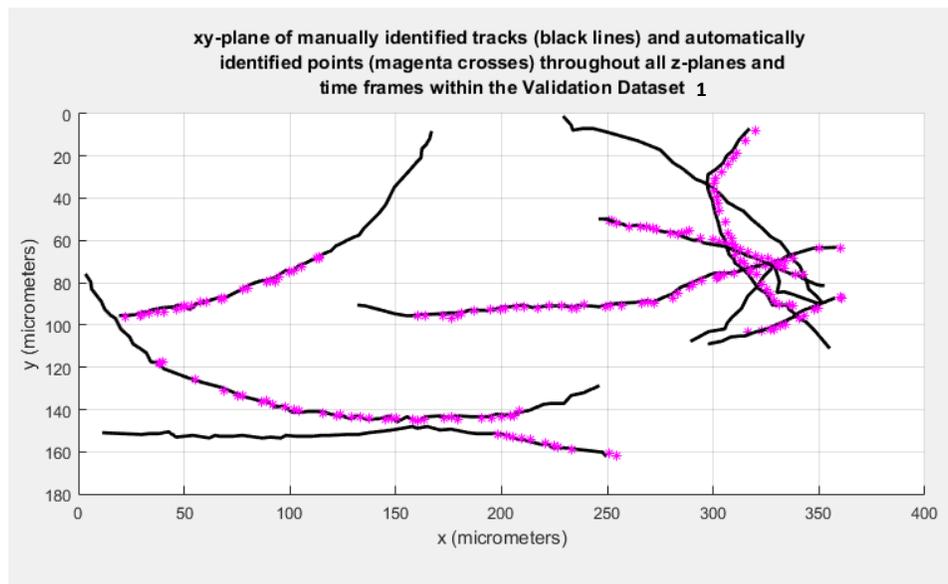


Figure 26
Top view of manually identified tracks (black) and automatically labelled points (magenta) throughout all the z planes and time frames Validation Dataset 1.

F_{0.5} Validation Score

The $F_{0.5}$ validation score is chosen to evaluate the performance of this algorithm because it places more emphasis on Precision rather than Recall (see section 3.2.3 for more details).

Table 10
Validation metrics of the microorganism identification algorithm using Validation Dataset 1

Total number of Points in Dataset	324
Total number of Points found by identification algorithm	187
True Positives	185
False Positives	2
False Negatives	139
Precision (P)	98.9%
Recall (R)	57.1%

$$F_{0.5} = (1 + 0.5^2) \cdot \frac{P \cdot R}{(0.5^2 \cdot P) + R} = (1 + 0.5^2) \cdot \frac{0.989 \cdot 0.571}{(0.5^2 \cdot 0.989) + 0.571}$$

$$F_{0.5} = 0.863 \tag{4.1}$$

Identification Error

The microorganisms within Validation Dataset 1 are *B. subtilis*. Depending on where it is in its cell cycle, *B. subtilis* ranges in from 3–9 μm in length and 0.5–1 μm in diameter [96, 97].

The identification error is defined as the three dimensional Euclidean distance between a microorganism's centroid found using the identification algorithm, and the actual location of that centroid found in the validation dataset. The sample dataset volume of validation dataset 1 is 360 μm x 360 μm x 252 μm .

The root-mean-squared (RMS) identification error for this dataset was found to be 7.42 μm .

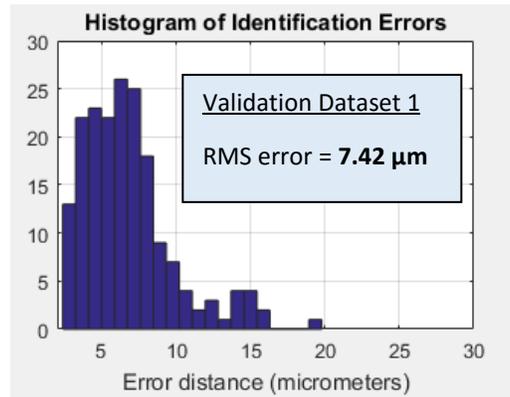


Figure 27

Histogram of error distances between manually labelled microorganisms and the microorganisms found using the identification algorithm in Validation Dataset 1. The RMS error distance is 7.42 μm .

PARTICLE IDENTIFICATION RESULTS FROM VALIDATION DATASET 2

Overview

The algorithm was trained using the validation dataset 2 volume at time-frame $t = 9$. This volume contained 6 *Colwellia psychrerythraea* microorganisms. Figure 24 displays a single z-slice of this training volume, which contains 2 of the 6 microorganisms in the training dataset.

The weight vector w was calculated using the feature matrix X_{train} and the user-labelled binary identification vector y_{train} . Parameters $pCutoff$ and $minCluster$ were tuned to identify all groups of microorganism pixels within this sample volume.

Table 11
Parameter values assigned using Training dataset of Validation Dataset 2.

Parameter name	Assigned value
$pCutoff$	0.03
$minCluster$	10

The calculated weight vector w was used to classify all the pixels in the entire validation dataset 2, and find the microorganisms present at all time frames. The results of the microorganism identification algorithm are presented in Figure 28 as a three-dimensional plot, and in Figure 29 as a top-view two-dimensional plot.

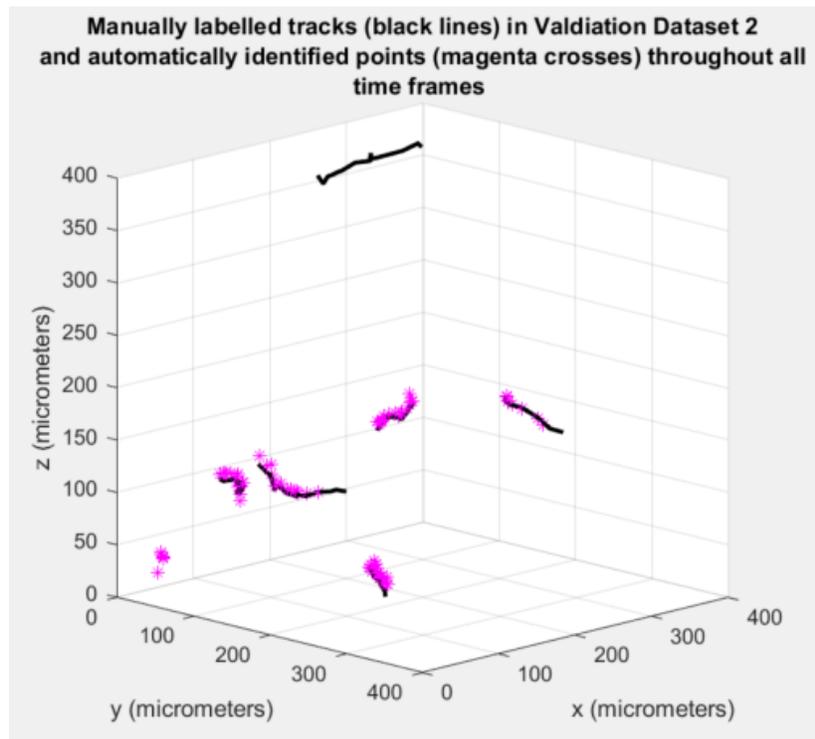


Figure 28
3D plot of manually identified tracks (black) and automatically identified points (magenta) throughout all time frames within Validation Dataset 2.

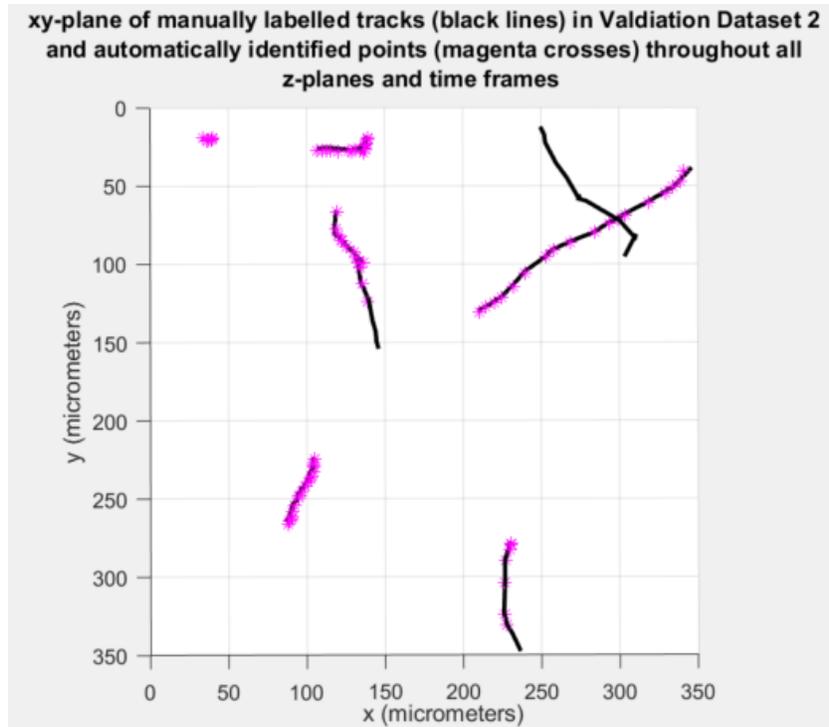


Figure 29

Top view of manually identified tracks (black) and automatically labelled points (magenta) throughout all the z planes and time frames Validation Dataset 2.

F_{0.5} Validation Score

The $F_{0.5}$ validation score is chosen to evaluate the performance of this algorithm because it places more emphasis on Precision than Recall (see section 3.2.3 for more details).

Table 12

Validation metrics of the microorganism identification algorithm using Validation Dataset 2

Total number of Points in Validation Dataset 2	98
Total number of Points found by identification algorithm	82
True Positives	75
False Positives	7
False Negatives	23
Precision (P)	91.5%
Recall (R)	76.5%

$$F_{0.5} = (1 + 0.5^2) \cdot \frac{P \cdot R}{(0.5^2 \cdot P) + R} = (1 + 0.5^2) \cdot \frac{0.915 \cdot 0.765}{(0.5^2 \cdot 0.915) + 0.765}$$

$$F_{0.5} = 0.880 \tag{4.2}$$

Identification Error

The microorganisms within validation dataset 2 are *Colwellia psychrerythraea*. They range in length from 2–4 μm and an average diameter of 1 μm [98].

The identification error is defined as the three dimensional Euclidean distance between a microorganism’s centroid found using the identification algorithm, and the actual location of that centroid found in the validation dataset. The sample volume of validation dataset 2 is 360 μm x 360 μm x 392 μm.

The RMS identification error in this dataset was found to be 5.59 μm (see Figure 30).

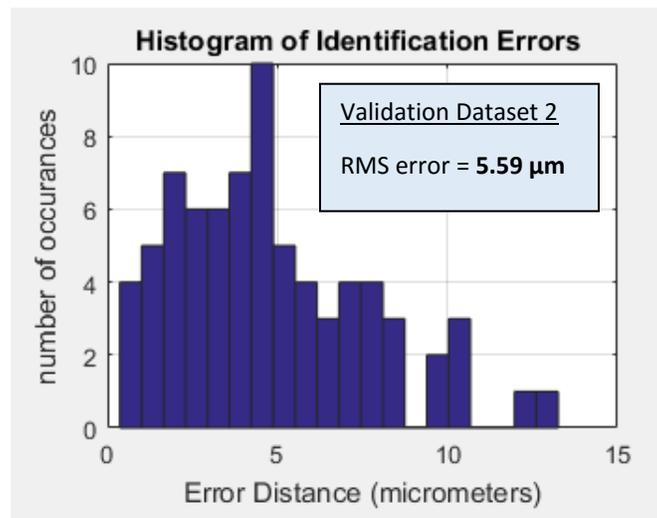


Figure 30

Histogram of error distances in Validation Dataset 2 between manually labelled microorganisms and the microorganisms found using the identification algorithm. The RMS identification error distance is 5.59 μm

4.2 Particle Tracking Results

The performance of any particle tracking algorithm depends heavily on the results of the particle identification algorithm used; if a particle cannot be identified, tracking it will be extremely challenging.

The x,y,z,t coordinates of the microorganisms identified in section 4.1 were used as inputs to a simple particle tracking algorithm based on nearest neighbour linking [91]. The particle tracking results of both validation datasets 1 and 2 are presented in Figure 31, Figure 32, Figure 33, and Figure 34.

The tracking algorithm performed as expected. It successfully linked the points found using the identification algorithm. There was no overlap in the automatically identified tracks between different microorganisms.

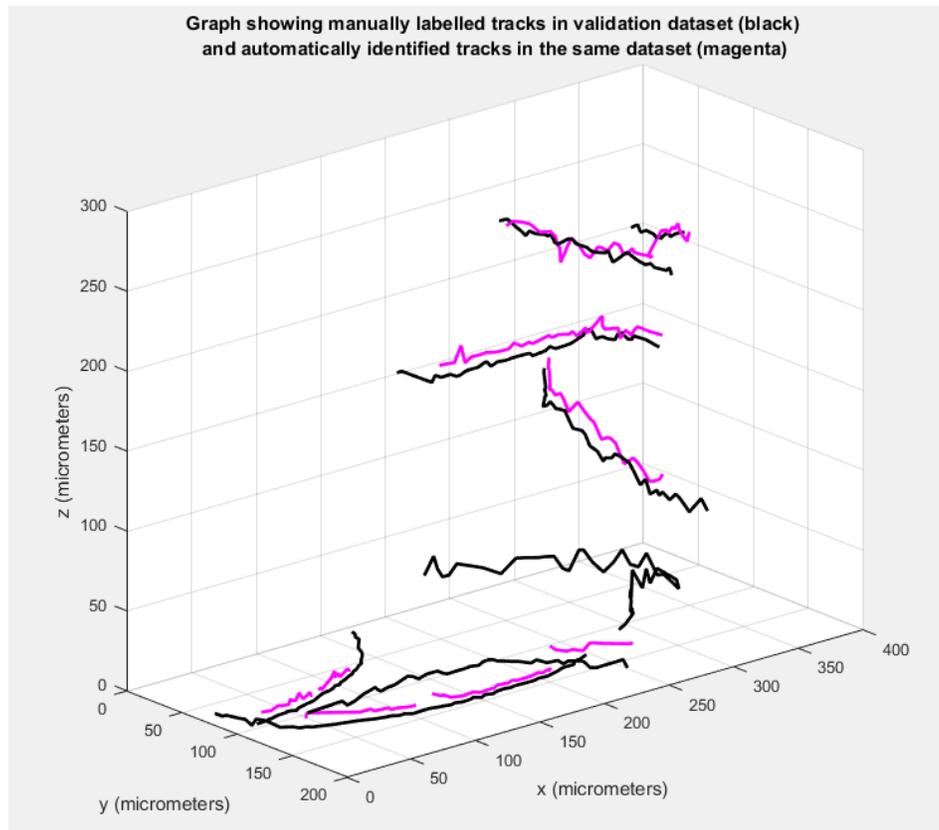


Figure 31
3D plot of manually labelled tracks (black) and automatically identified tracks (magenta) throughout all time frames in Validation Dataset 1

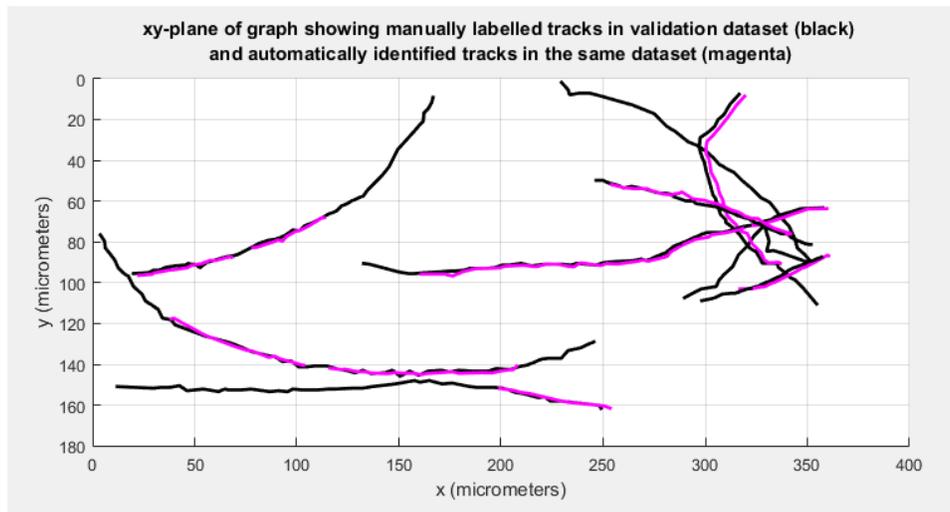


Figure 32
top view of manually identified tracks (black) and automatically labelled tracks (magenta) throughout all the z planes and time frames of Validation dataset 1.

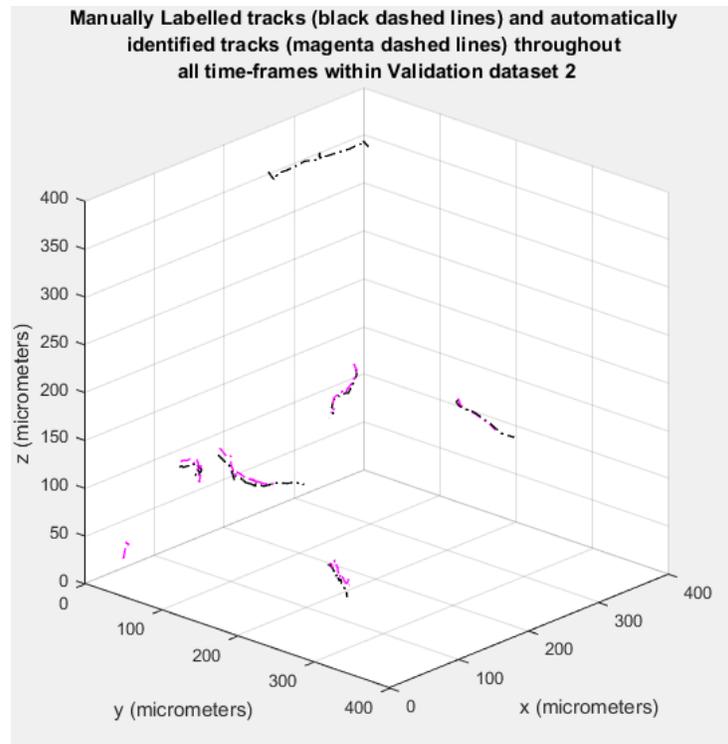


Figure 33

3D plot of manually labelled tracks (black dashed lines) and automatically labelled tracks (magenta dashed lines) throughout all time frames in Validation Dataset 2. The algorithm successfully found 5 out of 6 microorganism trajectories.

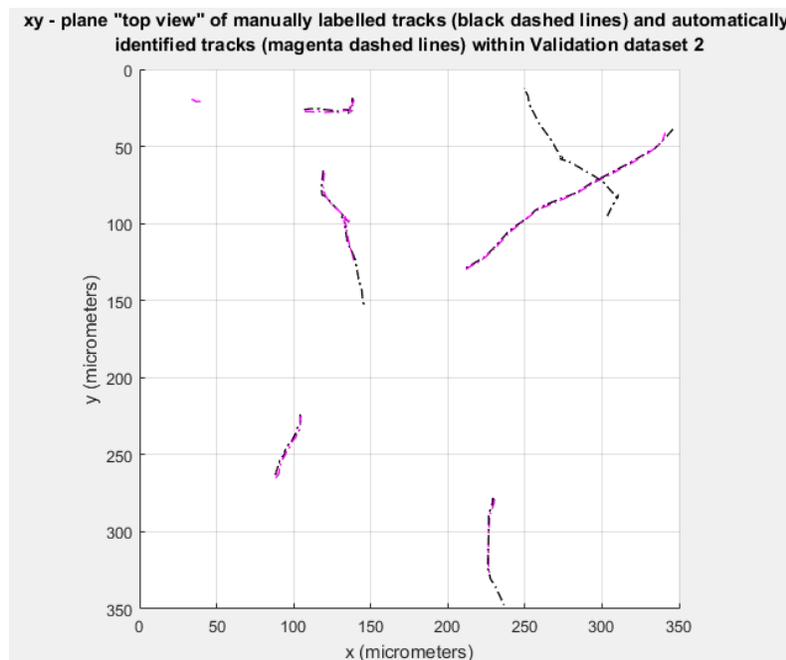


Figure 34

Top view of manually labelled tracks (black dashed lines) and automatically labelled tracks (magenta dashed lines) throughout all the z planes and time frames of Validation Dataset 2. The algorithm successfully found 5 out of 6 microorganism trajectories.

4.3 Properties of Microbial Motility

SWIMMING DIRECTION

The azimuthal angle is defined as the angle between the positive x axis and the direction vector of microbial motility projected onto the xy-plane. The azimuthal angle varies from $-\pi$ to π .

The elevation angle is defined as the angle between the xy-plane and the direction vector of microbial motility. It varies from $-\pi/2$ to $\pi/2$.

The majority of the swimming direction in Validation Dataset 1 was along the x-axis at all elevations and azimuth 0 , $-\pi$ and π (Figure 35). This result is consistent with the plotted tracks in the validation dataset (Figure 32), which shows most of the motility occurring in the x direction. There is a peak at direction (azimuth = 0 rad, elevation = 0 rad). This peak suggests that there might have been slight drift within the sample chamber in that direction.

In Validation Dataset 2, there is a peak (Figure 36) at direction (azimuth = 2.74 rad, elevation = 0 rad), suggesting that there might have been slight drift within the sample chamber in that direction.

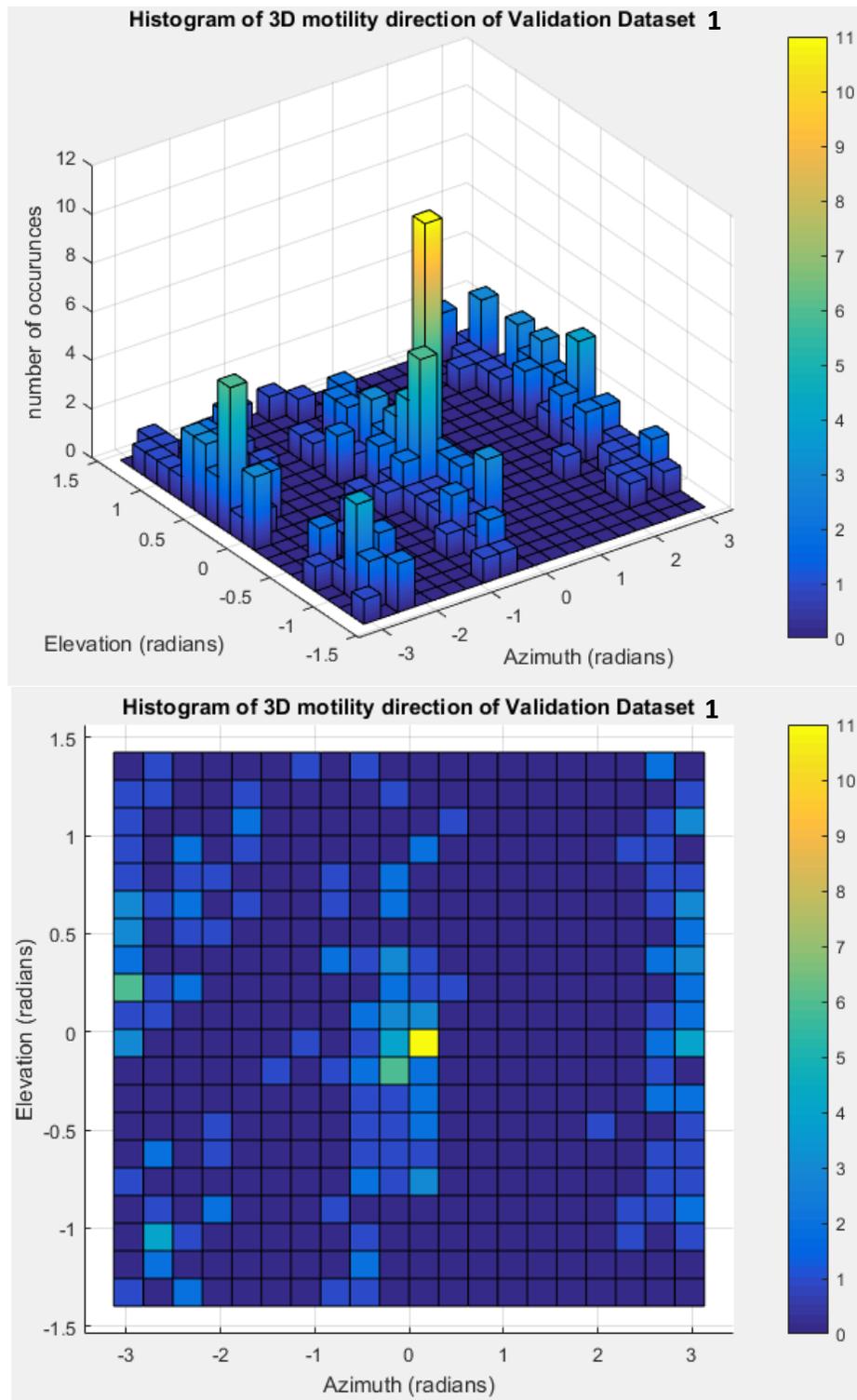


Figure 35
Histogram of the Instantaneous motility direction of the tracks found in Validation Dataset 1 using the identification algorithm

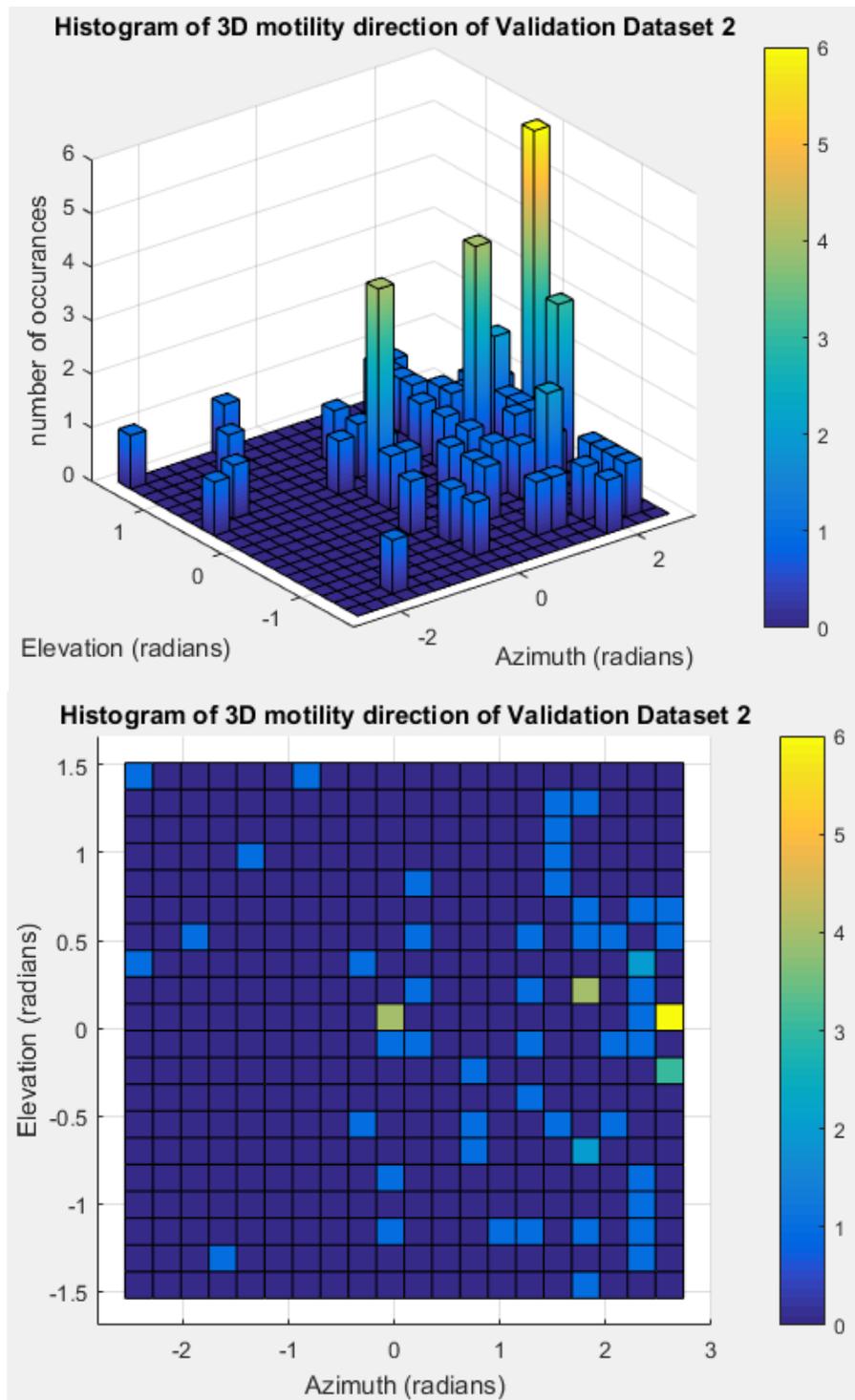


Figure 36
Histogram of instantaneous motility direction of the tracks found within Validation Dataset 2, using the identification algorithm.

SWIMMING SPEED

Ito et al. [99] measured the average swimming speed in 2D for *B. subtilis*, and found it to be 23 $\mu\text{m/s}$. In order to compare this with the results outlined in this study of Validation Dataset 1, the projection of the 3D tracks onto the xy plane was obtained and the 2D motility speeds were measured from these projections. A histogram of the 2D swimming speeds automatically identified by the algorithm is presented in Figure 37a; the median of the 2D swimming speeds was found to be 21.7 $\mu\text{m/s}$ with a wide range of variation. These swimming speeds are comparable with the literature.

On the other hand, the histogram of the 3D swimming speeds of the same dataset of microorganisms identified by the algorithm is presented in Figure 37b. It is observed that the median of the 3D swimming speeds is 28.8 $\mu\text{m/s}$.

Further investigation into the swimming speeds was conducted. The instantaneous 3D speeds of every track in validation dataset 1 were measured manually using user labelled locations of the cells at every time frame. These measurements were projected onto the xy-plane to generate 2D speeds, and then plotted against each time frame within the dataset (Figure 38). Figure 38 displays distinct peaks in the instantaneous speed at 9 out of 84 time frames (Black arrows). These peaks are spikes in the instantaneous speeds of at least double the value of both the previous and the next time frames. The peaks occur for every microorganism's tracks at that time frame.

This anomaly suggests that the time stamps recorded at those time frames are not exactly precise; the microorganisms were imaged a while after the timestamp, allowing them to move

further and therefore, to appear as though their instantaneous speeds have dramatically increased.

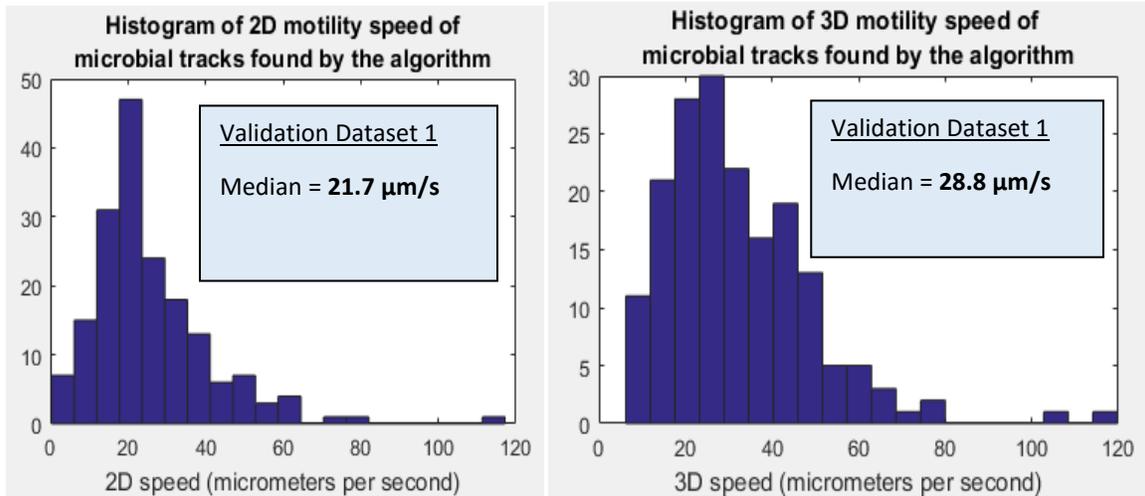


Figure 37(a, b)

Histogram of calculated microorganism speeds in Validation Dataset 1. The graphs show the same speeds. The 3D motility speed is calculated from the 3D tracks found using the algorithm. The 2D motility speed is calculated as a projection of the tracks from the z plane onto the xy axis, effectively generating a "Top View" of the microbial motility.

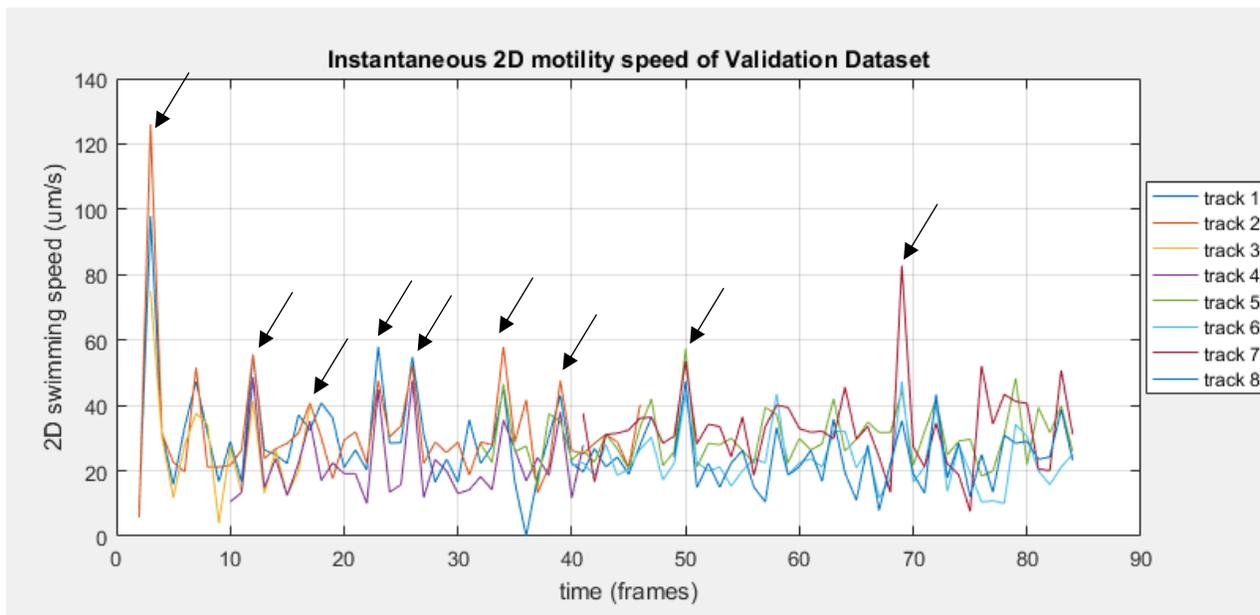


Figure 38

Manually User Labelled Instantaneous 2D speed of each track within the dataset. Black Arrows point to peaks in the instantaneous speeds for all tracks, suggesting that the time stamps recorded for these frames are erroneous.

CHAPTER 5: DISCUSSION, FUTURE WORK AND CONCLUSIONS

5.1 Discussion

Current techniques for observing bacterial motility are essentially two-dimensional because of the small depth of field provided by high numerical aperture microscope objectives. Measurement of 3D trajectories is inaccurate, since it is performed by approximating the third dimension from measured 2D trajectories, or by inferring the organisms' z positions as they travel into and out of focus. This means that for most bacteria, their swimming patterns in an unconstrained 3D volume still remain unknown. DHM is an emerging technology that can capture 3D trajectories of microorganisms. DHM phase images are used to create contrast between transparent bacteria and the medium that they swim in. One of the challenges of using DHM is the low signal-to-noise ratio (SNR) of the images captured, making automated identification and tracking the bacteria difficult using standard image processing and computer vision techniques. This thesis presents and validates a machine-learning identification method based on linear logistic regression that can consistently identify over 50% of all microorganisms within noisy DHM phase image reconstructions, with a precision of over 90%. Identification was validated using two different species of microorganisms of different sizes, without using any chemical contrast enhancement. Figure 31, Figure 32, Figure 33 and Figure 34 show the tracks obtained by using the algorithm, compared to the same tracks manually labelled by a human expert. Figure 13 demonstrates the performance of the algorithm on a third species of bacteria (*Vibrio alginolyticus*), and plots their 3D tracks. The identification algorithm presented in this thesis is demonstrated as flexible, simple and easy to implement.

The observed root-mean-squared identification error of the algorithm is nearly equal to the length of the microorganism being identified: 5.59 μm for the *Colwellia psychrerythraea*, and 7.42 μm for the *Bacillus subtilis* datasets. Tracking the cells through time was easy in the datasets shown here because of the low density of cells within the sample chamber. In higher density situations, a more sophisticated tracking algorithm may be required to account for overlap. However, one major advantage of 3D identification is that overlap incidents are much less frequent than in traditional 2D cases. This is because in 3D the cells must actually come into close contact with each other for an overlap to occur. On the other hand, in 2D, one cell must only pass in front of another for an overlap to occur regardless of how distant they actually are from each other.

This is the first time that Logistic Regression has been used to classify 3D holography datasets, segmenting the objects of interest from the background. It was found that extracting just a few simple 3D image pixel features was sufficient to obtain accurate results from the classifier. The most memory-intensive part of the algorithm was calculating the linear weights vector of the classifier using gradient descent regression (equation 3.2). The calculation for finding the weights vector for 9 features required having a computer configured with 32GB of random-access-memory (RAM).

The results obtained from the identification algorithm were used to calculate the microorganisms' instantaneous 3D swimming directions, and presented in Figure 35 and Figure 36. These histograms could be useful for detecting the direction of flow in 3D within the sample chamber, as well as for determining the 3D direction of taxis which may help observing bacterial swarming mechanisms.

The algorithm's results were also used to calculate the 3D swimming speeds of *Bacillus subtilis*, presented in Figure 37, and compared with those found in literature [99]. Since the literature only describes its 2D swimming speed, the 3D results obtained by the algorithm were projected onto the xy-plane to generate 2D speeds. The results were consistent with the literature. Comparing the 3D and its projected 2D speeds obtained in this thesis, it was found that the median of the 2D speeds was ~25% less than the median of the 3D speeds. This finding is comparable to the results presented by Taute et al. [17], who found the systematic errors between measuring 2D speeds of microorganisms versus their actual 3D speeds (see Figure 4).

When further investigation into the instantaneous swimming speeds was conducted and plotted in Figure 38, distinct peaks in the instantaneous speed at 9 out of 84 time frames were found. These peaks were spikes in the instantaneous speeds of at least double the value of both the previous and the next time frames, and occurred for every microorganism at that time frame. This anomaly suggests that maybe the time stamps recorded at those time frames were not precise; all the microorganisms were imaged a while after the timestamp, allowing them to move further and therefore, to appear as though their instantaneous speeds had dramatically increased. Therefore, more investigation into the accuracy of the timestamps file needs to be done before conducting any further speed-related experiments.

5.2 Future Work

INVESTIGATE THE TIME STAMPS

The accuracy of the time stamps file generated by the KOALA[®] software [95] should be investigated before conducting further experiments that depend on time, such as measurements

of speed. This could be done by imaging a constant flow of micro-particles using the DHM. The micro-particles' measured speeds, obtained by dividing their displacements over the time difference between the two time stamps, should be constant in this scenario.

Inaccuracies in the hologram time stamps could arise from slow writing of the data obtained by the DHM onto the PC hard disk. The DHM CCD is set to capture 4Mpx images at a rate of 15 fps, which is a very large data transfer rate. One suggestion would be to investigate if lowering the image resolution setting to 1Mpx could have a beneficial impact on the accuracy of the time stamps. If the hologram's image resolution is lowered, then the DHM's lateral resolution must again be calculated to ensure that the tradeoff in image quality does not affect the microscope's ability to resolve sub-micron scale structures.

INVESTIGATE DIFFERENT 3D FEATURES OF THE ALGORITHM

The algorithm presented in this thesis extracted 9 features to perform the classification. Each added feature increases the computational time required to complete the classification, and may increase the accuracy. However, increased accuracy is not guaranteed. Therefore, it would be interesting to find out which features have the most impact, and which have little to no impact. Furthermore, it would also be interesting to explore new features and their effect on improving the classification performance, whether these improvements increase the computational speed, or increase the accuracy of the classification, or both.

It may also be beneficial to investigate the effect of using a different machine-learning algorithm, such as convolutional neural networks.

INVESTIGATE THE EFFECT OF MICROBIAL DENSITY ON ALGORITHM PERFORMANCE

As bacterial samples get denser, all tracking algorithms have a harder time simultaneously tracking each and every cell within the sample. Further investigation on the impact of microbial density on the identification algorithm presented in this thesis would be interesting. However, this is difficult because validating the algorithm's performance requires an expert human to manually label all the microorganisms within the sample. This human step is very time consuming and tedious, especially for dense bacterial samples.

To avoid this issue, virtual data could instead be simulated to evaluate the effect of density on the algorithm's performance. However, this procedure would not take into account the lower SNR that will arise because of the multiple interference patterns created by the dense sample in the DHM.

CREATE A MOTILITY DATABASE

Thanks to the ease with which DHM microbial trajectories are obtained by using the algorithm in this thesis, an open source, 3D motility database can now be easily populated with motility properties of different microorganisms, properties such their respective swimming speeds and turning angles/mechanisms, in many different environments. A motility database may help us answer questions such as:

- i. What role does microbial motility play in aquatic environments: oceans, lakes, and inside the pores of sea ice?

- ii. When microbes lose the ability to swim, what will make them swim again--temperature, amino acids, sugars, or a combination of any these? How long does it take to go from motile to non-motile?
- iii. Is microbial motility substantially different in extremely cold and nutrient poor environments (such as the Arctic) than in more temperate environments?
- iv. How can motility be a biosignature that complements biochemical biosignatures of life on other planets?

5.3 Conclusion

In conclusion, this thesis presents a machine-learning approach to object identification within a DHM dataset. Machine-learning techniques are becoming much more prevalent now thanks to a massive proliferation of computing power within the last few decades. The identification algorithm presented in this thesis was able to accurately locate unstained microorganisms within a variety of noisy DHM phase image reconstructions, by adapting and “learning” the properties of the objects of interest. It is simple to implement and does not require explicit programming, making it flexible for a large variety of different types of datasets. Further development into learning algorithms seems promising, and these types of algorithms are expected to be found more frequently and within multiple disciplines in the near future.

REFERENCES

1. Qian, Y., M. Ye, and J. Zhou, *Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features*. IEEE Transactions on Geoscience and Remote Sensing, 2013. **51**(4): p. 2276-2291.
2. Ren, X. and J. Malik. *Learning a classification model for segmentation*. in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. 2003. IEEE.
3. Son, K., D.R. Brumley, and R. Stocker, *Live from under the lens: exploring microbial motility with dynamic imaging and microfluidics*. Nat Rev Micro, 2015. **13**(12): p. 761-775.
4. Berg, H.C. and R.A. Anderson, *Bacteria swim by rotating their flagellar filaments*. 1973.
5. Silverman, M. and M. Simon, *Flagellar rotation and the mechanism of bacterial motility*. Nature, 1974. **249**(5452): p. 73-74.
6. Macnab, R.M., *Bacterial flagella rotating in bundles: a study in helical geometry*. Proceedings of the National Academy of Sciences, 1977. **74**(1): p. 221-225.
7. Turner, L., W.S. Ryu, and H.C. Berg, *Real-time imaging of fluorescent flagellar filaments*. Journal of bacteriology, 2000. **182**(10): p. 2793-2801.
8. Taylor, B.L. and D. Koshland, *Reversal of flagellar rotation in monotrichous and peritrichous bacteria: generation of changes in direction*. Journal of bacteriology, 1974. **119**(2): p. 640-642.
9. Magariyama, Y., et al., *Difference in Bacterial Motion between Forward and Backward Swimming Caused by the Wall Effect*. Biophysical Journal, 2005. **88**(5): p. 3648-3658.
10. Goto, T., et al., *Comparison between observation and boundary element analysis of bacterium swimming motion*. JSME International Journal Series C, 2001. **44**(4): p. 958-963.
11. Xie, L., et al., *Bacterial flagellum as a propeller and as a rudder for efficient chemotaxis*. Proceedings of the National Academy of Sciences, 2011. **108**(6): p. 2246-2251.
12. Lane, N., *The unseen world: reflections on Leeuwenhoek (1677) 'Concerning little animals'*. Philosophical Transactions of the Royal Society B: Biological Sciences, 2015. **370**(1666).
13. Macnab, R.M., *The bacterial flagellum: reversible rotary propellor and type III export apparatus*. Journal of bacteriology, 1999. **181**(23): p. 7149-7153.
14. Nadeau, J., et al., *Microbial morphology and motility as biosignatures for outer planet missions*. Astrobiology, 2016. **16**(10): p. 755-774.
15. Saragosti, J., et al., *Directional persistence of chemotactic bacteria in a traveling concentration wave*. Proceedings of the National Academy of Sciences, 2011. **108**(39): p. 16235-16240.
16. Bubendorfer, S., et al., *Secondary bacterial flagellar system improves bacterial spreading by increasing the directional persistence of swimming*. Proceedings of the National Academy of Sciences, 2014. **111**(31): p. 11485-11490.
17. Taute, K., et al., *High-throughput 3D tracking of bacteria on a standard phase contrast microscope*. Nature communications, 2015. **6**.
18. Berg, H.C., *How to track bacteria*. Review of Scientific Instruments, 1971. **42**(6): p. 868-871.

19. Speidel, M., A. Jonáš, and E.-L. Florin, *Three-dimensional tracking of fluorescent nanoparticles with subnanometer precision by use of off-focus imaging*. Optics Letters, 2003. **28**(2): p. 69-71.
20. Wu, M., et al., *Collective bacterial dynamics revealed using a three-dimensional population-scale defocused particle tracking technique*. Applied and environmental microbiology, 2006. **72**(7): p. 4987-4994.
21. Li, G. and J.X. Tang, *Accumulation of microswimmers near a surface mediated by collision and rotational Brownian motion*. Physical review letters, 2009. **103**(7): p. 078101.
22. Edwards, M.R., et al., *Swimming characterization of Serratia marcescens for bio-hybrid micro-robotics*. Journal of Micro-Bio Robotics, 2014. **9**(3-4): p. 47-60.
23. Gabor, D., *A new microscopic principle*. Nature, 1948. **161**(4098): p. 777-778.
24. Gabor, D. *Microscopy by reconstructed wave-fronts*. in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. 1949. The Royal Society.
25. Schnars, U. and W.P. Jüptner, *Digital recording and numerical reconstruction of holograms*. Measurement science and technology, 2002. **13**(9): p. R85.
26. Powell, R.L. and K.A. Stetson, *Interferometric vibration analysis by wavefront reconstruction*. JOSA, 1965. **55**(12): p. 1593-1598.
27. Stetson, K.A. and R.L. Powell, *Interferometric hologram evaluation and real-time vibration analysis of diffuse objects*. JOSA, 1965. **55**(12): p. 1694-1695.
28. Kim, M.K., *Principles and techniques of digital holographic microscopy*. Journal of Photonics for Energy, 2010: p. 018005-018005-50.
29. Schnars, U. and W. Jüptner, *Direct recording of holograms by a CCD target and numerical reconstruction*. Applied Optics, 1994. **33**(2): p. 179-181.
30. Cuche, E., F. Bevilacqua, and C. Depeursinge, *Digital holography for quantitative phase-contrast imaging*. Optics Letters, 1999. **24**(5): p. 291-293.
31. Yaroslavskii, L. and N. Merzlyakov, *Methods of Digital Holography*.
32. Marquet, P., et al., *Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy*. Optics Letters, 2005. **30**(5): p. 468-470.
33. Chengala, A., M. Hondzo, and J. Sheng, *Microalga propels along vorticity direction in a shear flow*. Physical Review E, 2013. **87**(5): p. 052704.
34. Sheng, J., et al., *Digital holographic microscopy reveals prey-induced changes in swimming behavior of predatory dinoflagellates*. Proceedings of the National Academy of Sciences, 2007. **104**(44): p. 17512-17517.
35. Sheng, J., et al., *A dinoflagellate exploits toxins to immobilize prey prior to ingestion*. Proceedings of the National Academy of Sciences, 2010. **107**(5): p. 2082-2087.
36. Vater, S.M., et al., *Holographic microscopy provides new insights into the settlement of zoospores of the green alga Ulva linza on cationic oligopeptide surfaces*. Biofouling, 2015. **31**(2): p. 229-239.
37. Pavillon, N., et al., *Early cell death detection with digital holographic microscopy*. PloS one, 2012. **7**(1): p. e30912.

38. Kemper, B., et al., *Investigation of living pancreas tumor cells by digital holographic microscopy*. Journal of biomedical optics, 2006. **11**(3): p. 034005-034005-8.
39. Jourdain, P., et al., *Determination of transmembrane water fluxes in neurons elicited by glutamate ionotropic receptors and by the cotransporters KCC2 and NKCC1: a digital holographic microscopy study*. The Journal of Neuroscience, 2011. **31**(33): p. 11846-11854.
40. Janeckova, H., P. Vesely, and R. Chmelik, *Proving tumour cells by acute nutritional/energy deprivation as a survival threat: a task for microscopy*. Anticancer research, 2009. **29**(6): p. 2339-2345.
41. Miniotis, M.F., A. Mukwaya, and A.G. Wingren, *Digital holographic microscopy for non-invasive monitoring of cell cycle arrest in L929 cells*. PloS one, 2014. **9**(9): p. e106546.
42. Su, T.-W., L. Xue, and A. Ozcan, *High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories*. Proceedings of the National Academy of Sciences, 2012. **109**(40): p. 16018-16022.
43. Wyatt, P.J., *Differential light scattering: a physical method for identifying living bacterial cells*. Applied optics, 1968. **7**(10): p. 1879-1896.
44. Charrière, F., et al., *Shot-noise influence on the reconstructed phase image signal-to-noise ratio in digital holographic microscopy*. Applied optics, 2006. **45**(29): p. 7667-7673.
45. Molaei, M. and J. Sheng, *Imaging bacterial 3D motion using digital in-line holographic microscopy and correlation-based de-noising algorithm*. Optics express, 2014. **22**(26): p. 32119-32137.
46. Molaei Mehdi, M. and J. Sheng, *Imaging bacterial 3D motion using digital in-line holographic microscopy and correlation-based de-noising algorithm*. Optics Express, 2014. **22**(26): p. 32119-37.
47. Saxton, M.J. and K. Jacobson, *Single-particle tracking: applications to membrane dynamics*. Annual review of biophysics and biomolecular structure, 1997. **26**(1): p. 373-399.
48. Berginski, M.E., et al., *High-resolution quantification of focal adhesion spatiotemporal dynamics in living cells*. PloS one, 2011. **6**(7): p. e22025.
49. Brandenburg, B. and X. Zhuang, *Virus trafficking—learning from single-virus tracking*. Nature Reviews Microbiology, 2007. **5**(3): p. 197-208.
50. Jandt, U. and A.-P. Zeng, *Modeling of intracellular transport and compartmentation*, in *Genomics and Systems Biology of Mammalian Cell Culture*. 2011, Springer. p. 221-249.
51. Sinha, B., et al., *Dynamic organization of chromatin assembly and transcription factories in living cells*. Methods in cell biology, 2010. **98**: p. 57-78.
52. Agarwal, S., et al., *ATP-dependent and independent functions of Rad54 in genome maintenance*. The Journal of cell biology, 2011. **192**(5): p. 735-750.
53. Meijering, E., O. Dzyubachyk, and I. Smal, *Methods for cell and particle tracking*. Methods Enzymol, 2012. **504**(9): p. 183-200.
54. Meijering, E., I. Smal, and G. Danuser, *Tracking in molecular bioimaging*. IEEE signal processing magazine, 2006. **23**(3): p. 46-53.

55. Kalaidzidis, Y., *Intracellular objects tracking*. European journal of cell biology, 2007. **86**(9): p. 569-578.
56. Dorn, J.F., G. Danuser, and G. Yang, *Computational processing and analysis of dynamic fluorescence image data*. Methods in cell biology, 2008. **85**: p. 497-538.
57. Meijering, E., et al. *Tracking in cell and developmental biology*. in *Seminars in cell & developmental biology*. 2009. Elsevier.
58. Jaqaman, K. and G. Danuser, *Computational image analysis of cellular dynamics: a case study based on particle tracking*. Cold Spring Harbor Protocols, 2009. **2009**(12): p. pdb. top65.
59. Rohr, K., et al., *Tracking and quantitative analysis of dynamic movements of cells and particles*. Cold Spring Harbor Protocols, 2010. **2010**(6): p. pdb. top80.
60. Sonka, M., V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*. 2014: Cengage Learning.
61. Chenouard, N., et al., *Objective comparison of particle tracking methods*. Nature methods, 2014. **11**(3): p. 281.
62. Sbalzarini, I.F. and P. Koumoutsakos, *Feature point tracking and trajectory analysis for video imaging in cell biology*. Journal of structural biology, 2005. **151**(2): p. 182-195.
63. Coraluppi, S. and C. Carthel, *Recursive track fusion for multi-sensor surveillance*. Information Fusion, 2004. **5**(1): p. 23-33.
64. Coraluppi, S. and C. Carthel, *Multi-Stage Multiple-Hypothesis Tracking*. J. Adv. Inf. Fusion, 2011. **6**(1): p. 57-67.
65. Olivo-Marin, J.-C., *Extraction of spots in biological images using multiscale products*. Pattern recognition, 2002. **35**(9): p. 1989-1996.
66. Chenouard, N., I. Bloch, and J.-C. Olivo-Marin. *Multiple hypothesis tracking in cluttered condition*. in *ICIP*. 2009.
67. Chenouard, N., I. Bloch, and J.-C. Olivo-Marin. *Multiple hypothesis tracking in microscopy images*. in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2009. IEEE.
68. Winter, M.R., et al., *Axonal transport analysis using multitemporal association tracking*. International journal of computational biology and drug design, 2012. **5**(1): p. 35-48.
69. Winter, M., et al., *Vertebrate neural stem cell segmentation, tracking and lineaging with validation and editing*. Nature Protocols, 2011. **6**(12): p. 1942-1952.
70. Godinez, W.J., et al., *Deterministic and probabilistic approaches for tracking virus particles in time-lapse fluorescence microscopy image sequences*. Medical image analysis, 2009. **13**(2): p. 325-342.
71. Godinez, W.J., et al. *Tracking multiple particles in fluorescence microscopy images via probabilistic data association*. in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2011. IEEE.
72. Rink, J., et al., *Rab conversion as a mechanism of progression from early to late endosomes*. Cell, 2005. **122**(5): p. 735-749.

73. Liang, L., et al. *Tracking clathrin coated pits with a multiple hypothesis based method*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2010. Springer.
74. Yin, Z., T. Kanade, and M. Chen, *Understanding the phase contrast optics to restore artifact-free microscopy images for segmentation*. *Medical image analysis*, 2012. **16**(5): p. 1047-1062.
75. Magnusson, K.E. and J. Jaldén. *A batch algorithm using iterative application of the Viterbi algorithm to track cells and construct cell lineages*. in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. 2012. IEEE.
76. Husain, M., et al., *Software for drift compensation, particle tracking and particle analysis of high - speed atomic force microscopy image series*. *Journal of Molecular Recognition*, 2012. **25**(5): p. 292-298.
77. Casuso, I., et al., *Characterization of the motion of membrane proteins using high-speed atomic force microscopy*. *Nature nanotechnology*, 2012. **7**(8): p. 525-529.
78. Rao, A.R. and B.G. Schunck, *Computing oriented texture fields*. *CVGIP: Graphical Models and Image Processing*, 1991. **53**(2): p. 157-185.
79. Hager, G.D., M. Dewan, and C.V. Stewart. *Multiple kernel tracking with SSD*. in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. 2004. IEEE.
80. Rousseeuw, P.J. and A.M. Leroy, *Robust regression and outlier detection*. Vol. 589. 2005: John Wiley & Sons.
81. Thompson, R.E., D.R. Larson, and W.W. Webb, *Precise nanometer localization analysis for individual fluorescent probes*. *Biophysical journal*, 2002. **82**(5): p. 2775-2783.
82. Shafique, K. and M. Shah, *A noniterative greedy algorithm for multiframe point correspondence*. *IEEE transactions on pattern analysis and machine intelligence*, 2005. **27**(1): p. 51-65.
83. Jaqaman, K., et al., *Robust single-particle tracking in live-cell time-lapse sequences*. *Nature methods*, 2008. **5**(8): p. 695-702.
84. Lowe, D.G., *Distinctive image features from scale-invariant keypoints*. *International journal of computer vision*, 2004. **60**(2): p. 91-110.
85. Crocker, J.C. and D.G. Grier, *Methods of digital video microscopy for colloidal studies*. *Journal of colloid and interface science*, 1996. **179**(1): p. 298-310.
86. Celler, K., G.P. van Wezel, and J. Willemse, *Single particle tracking of dynamically localizing TatA complexes in Streptomyces coelicolor*. *Biochemical and biophysical research communications*, 2013. **438**(1): p. 38-42.
87. Ku, T.-C., et al., *An automated tracking system to measure the dynamic properties of vesicles in living cells*. *Microscopy research and technique*, 2007. **70**(2): p. 119.
88. Ku, T.C., et al., *Morphological filter improve the efficiency of automated tracking of secretory vesicles with various dynamic properties*. *Microscopy research and technique*, 2009. **72**(9): p. 639-649.
89. Sheng, J., E. Malkiel, and J. Katz, *Digital holographic microscope for measuring three-dimensional particle distributions and motions*. *Applied Optics*, 2006. **45**(16): p. 3893-3901.

90. Hosmer Jr, D.W. and S. Lemeshow, *Applied logistic regression*. 2004: John Wiley & Sons.
91. Tinevez, J.-Y. and Y. Cao, *Simple Tracker*. MATLAB Central File Exchange, 2016.
92. Wallace, J.K., et al., *Robust, compact implementation of an off-axis digital holographic microscope*. *Optics Express*, 2015. **23**(13): p. 17367-17378.
93. Wallace, J.K., et al., *A Robust, Common-Mode Digital Holographic Microscope*.
94. Lindensmith, C.A., et al., *A Submersible, Off-Axis Holographic Microscope for Detection of Microbial Motility and Morphology in Aqueous and Icy Environments*. *PloS one*, 2016. **11**(1): p. e0147700.
95. LynceeTec. *KOALA ACQUISITION & ANALYSIS*. Available from: <http://www.lynceetec.com/koala-acquisition-analysis/>.
96. Sharpe, M.E., et al., *Bacillus subtilis cell cycle as studied by fluorescence microscopy: constancy of cell length at initiation of DNA replication and evidence for active nucleoid partitioning*. *Journal of bacteriology*, 1998. **180**(3): p. 547-555.
97. Henriques, A.O., et al., *Control of cell shape and elongation by the rodA gene in Bacillus subtilis*. *Molecular microbiology*, 1998. **28**(2): p. 235-247.
98. Junge, K., H. Eicken, and J.W. Deming, *Motility of Colwellia psychrerythraea strain 34H at subzero temperatures*. *Applied and environmental microbiology*, 2003. **69**(7): p. 4282-4284.
99. Ito, M., et al., *Properties of Motility in Bacillus subtilis Powered by the H(+)- Coupled MotAB Flagellar Stator, Na(+)-Coupled MotPS or Hybrid Stators MotAS or MotPB*. *Journal of molecular biology*, 2005. **352**(2): p. 396-408.