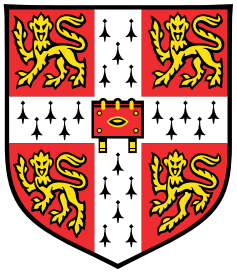
A Universal Image-Based Autofocus Device for Video Microscopy

Application to Non-Invasive Imaging of the Conjunctival Microcirculation with Single-Cell Resolution



Thomas Elliott Lynch

Department of Engineering
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Hughes Hall November 2025

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration or content generated by artificial intelligence, except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Thomas Elliott Lynch
November 2025

Abstract

A fundamental trade-off in microscopy is between resolution and depth-of-field (DOF). This limitation is a particular challenge in video microscopy, where axial motion of the sample, objective, or stage can rapidly cause defocus that cannot be corrected offline. Classical image-based autofocus methods "hunt" because the sign of defocus is unknown *a-priori*, precluding autofocus at video rate. Existing radiation-based autofocus devices fail when the imaging target lacks a stable relation to a coverslip surface. These constraints are especially limiting for slit-lamp microscopy of the conjunctival microcirculation, where microsaccades, blink dynamics, respiration, and heart action introduce axial motion that degrades focus consistently.

This thesis introduces a self-contained, portable, image-based autofocus device designed as a universal module that mounts on the camera port of any infinity-corrected microscope. The refocusing principle pairs a tilted sensor for depth-sensing with a tunable air-lens for refocusing. The tilted sensor's transverse sharpness profile provides both the sign and magnitude of defocus from a single frame. The air-lens then applies a corresponding change in optical power, by modulating the distance between its two elements.

A real-time control and acquisition software stack closes the sensing-to-actuation loop at video rate (60 Hz). A novel algorithm determines the LoBF on the tilted sensor in 1.34 ms, while remaining robust to sparse edge distribution, brightness differences, and specular reflection. A gain-scheduled proportional-derivative controller translates the LoBF stream to air-lens commands, continually adjusting to changing object distances. A C++ codebase achieves simultaneous operation of the cameras, LoBF algorithm, image registration, lens control, and GUI interaction, all on commercial hardware.

Performance characterization demonstrates that the controller achieves rapid, accurate convergence to the focus set-point in response to near-step disturbances. Settling times range from 50 ms to 142 ms in response to 0.1 mm to 1.0 mm perturbations, with no detectable steady-state error. Optical neutrality was verified by comparing a camera-only path with one including the autofocus optics: spatial resolution is preserved within 4-8% of reference, and depth-of-field is unaffected. Across the full autofocus range, image magnification is maintained to within 3.4%. The corresponding object-space working range is measured

from 0.5–71.25 mm (median 9.8 mm) across eleven objectives on three microscopes. These results validate an optically-neutral, high-speed autofocus system with a working range far exceeding existing commercial systems.

We demonstrate the capability of this autofocus device for *in-vivo*, non-invasive conjunctival imaging on a clinical slit-lamp. A paired-port experiment finds that the fraction of frames meeting the in-focus criterion was $88.0 \pm 8.8\%$ with autofocus versus $52.9 \pm 30.3\%$ under fixed focus. Supplementary Videos S1-S3 demonstrate that capillary loops, vessel bifurcations, and blood flow remain crisp, with single-erythrocyte resolution maintained. Routine access to focused video of the conjunctival microcirculation enables clinically-relevant biomarker extraction, yielding per-vessel velocities (V_a, V_c), diameters ($D(s)$), and derived haemodynamic quantities (Q, WSR, WSS). Beyond ophthalmic imaging, the same module maintains focus during polymer-sponge tensile tests on a benchtop stereomicroscope, and separately across a heterogeneous set of bespoke geometric graticules and real-world targets. These results confirm clinical value, and illustrate the device's ability to generalize to other microscope platforms and imaging applications.

In sum, this work (i) introduces and validates a universal image-based autofocus device for video microscopy, (ii) delivers a real-time control and acquisition software stack, (iii) demonstrates clinical-grade applicability to conjunctival imaging on a slit-lamp, and (iv) establishes universality beyond the slit-lamp.

List of approved additional materials

The following supplemental videos accompany this thesis, providing visual demonstrations of the system's performance. They are provided electronically for download on Zenodo [74], and with the physical thesis.

S1 Conjunctival imaging with autofocus: Video of the conjunctival microcirculation with active autofocus, demonstrating live focus-holding during imaging (40 s).

S2 Conjunctival imaging with autofocus (0.5×): Slow-motion video illustrating single-erythrocyte resolution and focus stability (33 s).

S3 Conjunctival imaging with and without autofocus: Simultaneous side-by-side recording showing improvement in image stability and clarity (18 s).

S4 Polymer sponge tensile test with and without autofocus: Side-by-side autofocus and fixed-focus views of a polymer sponge tensile test on a benchtop microscope, illustrating focus-holding and cross-platform generalization (25 s).

S5 Perturbed circuit board with and without autofocus: Side-by-side autofocus and fixed-focus views of a printed-circuit board undergoing dynamic perturbation, demonstrating generalization across imaging targets (9 s).

S6 Autofocus software interface: Screen recording of the autofocus software application, highlighting the depth-mapping and double-click-to-autofocus features (39 s).

Acknowledgements

This thesis is only possible thanks to the support, generosity, and good humour of many people.

First, I would like to thank Thierry Savin and Paul Meyer for their fantastic support as my supervisors. Paul was a wellspring of design and clinical expertise, and he continually pushed me to higher standards in my work and my writing. Thierry provided a balance of rigour and reassurance, guiding me toward meetings and conference stages that always proved more valuable than I anticipated. Between them, I learned to aim high and argue carefully.

I gratefully acknowledge the funding from the Engineering and Physical Sciences Research Council (EPSRC).

From Cambridge Enterprise, my sincere thanks to Margaret Wilkinson, Matt Wells, and Matt Cairns for their guidance and support in preparing and securing a patent for this device.

Several outstanding students infused the project with energy and fresh ideas. Clare Zhou, Eric Song, Will Holmes, Blendi Bylygbashi, Lohith Konathala, and Isaac Barden each contributed as part of their fourth-year undergraduate theses. I am also indebted to Sophia Constantinou for shepherding the project handover in my final few months.

I have been exceedingly lucky with my colleagues and communities. My crewmates and coaches at the Cambridge University Boat Club insisted on high standards, both athletic and academic, and gave me countless opportunities to explain my research. The Division C lunch group provided an ideal supply of willing test subjects. More importantly, they were a much-needed reprieve during long days in my windowless lab. The Cambridge University Entrepreneurship Society, Artificial Intelligence Society, and HomeDAO provided friends and career opportunities, both during and beyond my degree.

Finally, my family. To my brothers, Stephen, James, and Ben Lynch, for the regular reminders that I'm not as clever as I think; and to my parents, Kevin and Claire Lynch, for the equally regular reminders that I'm not so bad either.

Thank you all.

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

D	Vessel diameter
d	Separation of the air-lens elements
D_c	Erythrocyte (cell) diameter
e	Tilted sensor pixel-domain error signal, $e = x - x_0$
F	Generic sharpness score
f	Focal length of an optical component
f_o	Objective focal length
f_T	Tube-lens focal length
F_{TEN}	Sharpness score given by the Tenengrad measure
F_{VOL}	Sharpness score given by the Vollath F_4 measure
G_x, G_y	Image gradients following convolution with a sharpness operator
$I(i, j)$	Image intensity at pixel (i, j)
K_d	Derivative gain
K_i	Integral gain
K_p	Proportional gain
μ	Dynamic viscosity of blood
P	Optical power

p Pixel pitch of the tilted sensor (m·px^{-1})

P_{air} Optical power of the air-lens

P_T Optical power of the tube lens ($P_T = 1/f_T$)

Q Volumetric flow rate

S Sharpness image

θ Tilt angle of the tilted sensor

t_s Settling time

V_a Axial (centreline) flow velocity

V_c Cross-sectional flow velocity

x_0 LoBF coordinate at best focus / controller setpoint (pixels)

x LoBF on the tilted sensor (pixels)

x_C LoBF from a COM-based reduction (pixels)

x_G LoBF from a Gaussian-fit reduction (pixels)

x_{TG} LoBF from a Truncated-Gaussian reduction (pixels)

z_{im} Image-space location of the nominal image plane

z_{obj} Object-space axial location of the specimen

Acronyms / Abbreviations

AF Autofocus

AUROC Area Under the Receiver Operating Characteristic

CAD Computer-Aided Design

CCD Charge-Coupled Device

CLAHE Contrast-Limited Adaptive Histogram Equalization

CL Contact Lens

CMOS Complementary Metal-Oxide–Semiconductor

CNN Convolutional Neural Network

COM Centre-of-Mass

CPU Central Processing Unit

CVD Cardiovascular Disease

DED Dry Eye Disease

DNN Deep Neural Network

DOF Depth of Field

ECC Enhanced Correlation Coefficient

FFT Fast Fourier Transform

FOV Field of View

FPGA Field-Programmable Gate Array

FPS Frames per Second

FSLB Functional Slit-Lamp Biomicroscopy

FWHM Full-Width at Half-Maximum

GPU Graphics Processing Unit

GTK GIMP Toolkit (GUI library)

GUI Graphical User Interface

HE Histogram Equalization

HVI Hemoglobin Video Imaging

I/O Input/Output

IR Infrared

LK Lucas–Kanade

LoBF Location of Best Focus

MII Minimum-Intensity Image

MI Myocardial Infarction

NUC Next Unit of Computing (Intel mini-PC form factor)

PCB Printed-Circuit Board

PID Proportional–Integral–Derivative

QE Quantum Efficiency

RAM Random Access Memory

RBC Red Blood Cell

ROI Region of Interest

SDK Software Development Kit

SDL Simple DirectMedia Layer (graphics library)

SIFT Scale-Invariant Feature Transform

SL Slit Lamp

SoC System-on-Chip

STI Space–Time Image (for velocity estimation)

STIV Space–Time Image Velocimetry

UI User Interface

U-Net Convolutional Neural Network for Image Segmentation

USB Universal Serial Bus

VPP Velocity pulse period

WSI Whole-Slide Imaging

WSR Wall Shear Rate

WSS Wall Shear Stress

Chapter 1

Introduction

1.1 Motivation

A fundamental trade-off in microscopy is between resolution and depth-of-field (DOF), where higher resolution typically requires a shallow DOF. This is because higher numerical apertures, which improve resolution, also reduce the axial range over which features remain sharply focused. This trade-off is a particular challenge in video microscopy, where axial defocus from sample, objective, or stage motion can occur abruptly during acquisition and accumulate over time. Movement artefacts in the image plane, orthogonal to the optical axis, can be accommodated with image registration during- or post-acquisition [80, 29, 76]. In contrast, loss of focus cannot be recovered offline [123, 34]. Therefore, motion along the optical axis must be accommodated for during image acquisition.

To address this challenge, autofocus systems have been developed, which broadly fall into two categories: radiation-based and image-based. Image-based autofocus classically modulates focus, evaluates the difference in image quality using a sharpness metric, and then steps the optics toward an optimum [107, 24]. Because the axial direction of motion is unknown *a priori*, these methods rely on initial random changes in optical power and iterative adjustment toward the sharpness peak. Intermittent loss-of-focus is therefore inherent, precluding autofocus at video-rate [17].

Radiation-based systems, in contrast, typically reflect near-infrared radiation off a coverslip–sample interface to determine range, and then drive the objective or stage to maintain steady object distance [61, 111]. These systems fail when the imaging target lacks a stable relation to the coverslip surface, and they are incompatible when a coverslip or stage is unavailable. More generally, they require microscope-specific geometries where the objective or stage can be actuated, and often necessitate hardware modifications to the internal optical path [120].

These constraints are especially limiting for slit-lamp microscopy. A slit-lamp is a microscope used routinely in ophthalmology to examine the human eye *in vivo*. Unlike a conventional microscope, it has no coverslip or stage, and the imaged ocular surface is both curved and translucent. The tear film, conjunctiva, and episclera form a variable semi-reflective surface, so radiation-based sensors cannot lock onto a stable plane. As a result, both radiation-based and conventional image-based autofocus techniques are unsuited to maintaining continuous focus in slit-lamp imaging conditions.

The shortcomings of existing autofocus methods restrict valuable opportunities for research and diagnosis. The conjunctiva represents a particularly attractive site for repeatable, non-invasive assessment of the microcirculation [52, 72, 80, 112, 8]. Long, in-focus videos of the conjunctival microvasculature with single-erythrocyte resolution would enable reliable quantification of capillary hemodynamics over time, with relevance to both ocular and systemic disease [52]. However, microsaccades, blink dynamics, respiration, and heart action introduce axial motion that degrades focus consistently in practice [84, 8]. As a result, untrained clinicians and researchers are unable to non-invasively acquire high-resolution videos of the conjunctiva, leaving long-term monitoring of the microcirculation effectively inaccessible.

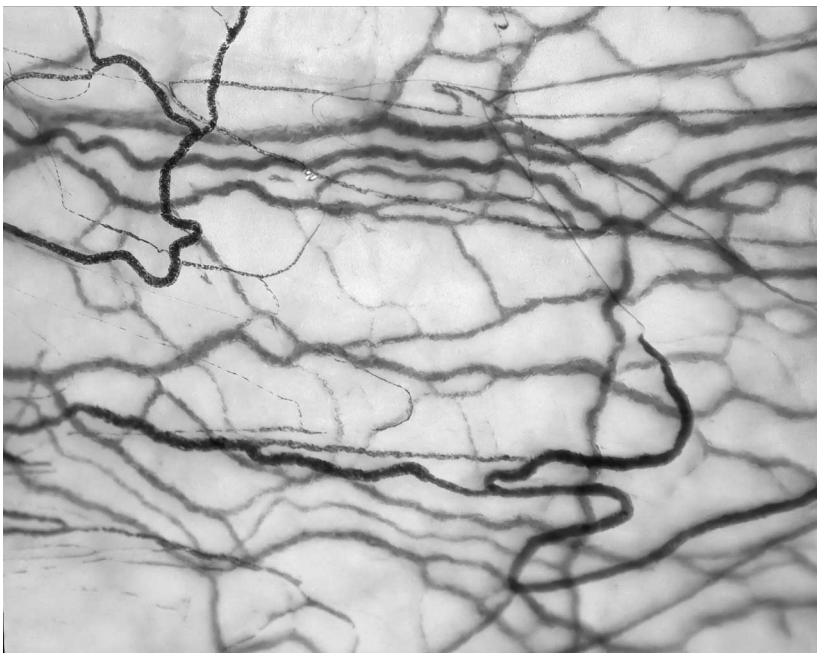


Fig. 1.1 Sample image of the conjunctival microcirculation, where each black dot is an individual erythrocyte. Supplementary Video S1 is taken using the autofocus device and demonstrates this microcirculation.

Accordingly, the primary motivation of this work is to allow untrained clinicians and researchers to reliably produce high-resolution videos of ocular microvascular blood flow for automated analysis. A secondary aim was to design the autofocus device as a universal module, generalizable to other microscopes and imaging modalities, so that its use would not be restricted to the slit-lamp. To achieve these aims, we introduce a self-contained, portable, image-based autofocus device that mounts on the camera port of any infinity-corrected microscope. Operating entirely within the collimated beam at the camera port, the device is designed to be objective-agnostic and deliver robust, video-rate focus stability for slit-lamp microscopy and beyond. We demonstrate the capability of this autofocus device by producing 30-second in-focus videos of the conjunctival microcirculation with single-cell resolution. Then, we demonstrate that the autofocus device generalises to bench-top microscopy, and to a variety of imaging targets beyond conjunctival imaging.

1.2 Prior work

At project start, the metrology lab had a Zeiss SL120 slit-lamp mounted on a stabilised bench for conjunctival imaging. The slit lamp was modified for Hemoglobin-Video Imaging (HVI), which selectively illuminates red blood cells using 505–575 nm (~green) filtered light. HVI is discussed further in Section 2.4.1.

An early prototype demonstrating a novel optomechanical principle for image-based autofocus was in place. Paul Meyer conceived and built this early prototype, which combined a tilted auxiliary sensor with an adjustable-focus air lens. The tilted sensor reported the location of best focus along its horizontal axis so that the direction and magnitude required to return focus to an imaging camera could be inferred from a single frame. The air lens then applied the corresponding change in optical-power. This refocusing principle is discussed further in Section 3.2.

Prototype control software existed, but was not viable for videography. In 2020, Claire Zhou implemented a Python pipeline as part of her fourth-year undergraduate thesis. This pipeline was able to update the focal plane up to a maximum rate of 5.1 Hz (n=3), which was insufficient for video-rate autofocus. As a result, videos taken using the prototype would lose focus in response to even slow specimen perturbations.

Accuracy problems were also substantial and recurrent:

- The algorithm often failed when one side of the image was brighter than the other. This was common, as the SL120 slit-lamp used an oblique light source.

- The algorithm could not manage specular reflections, which are common when illuminating the reflective conjunctiva.
- The algorithm depended on edges being well distributed across the image to determine the location of best focus. When edges were concentrated in only part of the field, the lens often settled incorrectly or oscillated between two out-of-focus positions.
- The system was too insensitive, failing to move the lens and correct the focal plane despite a slight loss of focus. It was also occasionally too sensitive, leading to oscillation and instability in the lens actuator.

For in-plane motion and basic acquisition, Tom Drummond in 2014 had written software to stream from a Prosilica GC1380H camera and perform real-time image stabilization. Drummond achieved real-time image stabilization by detecting edges based on large second derivatives of intensity, and tracking them via 1D searches perpendicular to the vessel to estimate global motion. This code base was old, remaining unopened since 2014, and it lacked any documentation. It also operated only on a MacBook whose operating system hadn't been updated since 2014. The Prosilica camera was a discontinued ethernet-based model, and difficult to integrate with contemporary libraries.

In summary, the lab provided a useful starting point: a slit lamp modified for HVI illumination, a promising tilted-sensor plus air-lens early prototype, and legacy software for image stabilization and camera playback. It was not adequate for continuous, in-focus videography. Ocular microsaccades, blinks, respiration, and cardiac motion require higher-rate axial control, robustness to brightness asymmetry and specular reflections, and a maintainable software and control stack [84]. These gaps motivate the software application, control architecture, and video-rate, image-based autofocus device developed in this thesis.

1.3 Approach

Building on the motivations and the inherited platform described above, my approach was to:

(i) validate the problem–solution fit for a tilted-sensor plus tunable air-lens architecture;
(ii) engineer a universal device around that concept;
(iii) design a software and control stack for accessible and reliable video-rate operation;
(iv) demonstrate applicability to conjunctival imaging and automated analysis of blood flow through the microcirculation; and
(v) demonstrate generalizability to other microscopes and imaging targets.

Development proceeded iteratively, with hardware and software refined together over successive prototypes. For clarity, however, the thesis is organised to follow the maturation of the core ideas.

In Chapter 2, I review autofocus strategies in microscopy and show how the tilted-sensor plus tunable air-lens architecture emerges as a viable approach for conjunctival imaging and for video microscopy in general. In parallel, I review conjunctival imaging techniques and their clinical value, highlighting key biomarkers that can be extracted from focused videos. This work sets the design criteria for image-based operation and video-rate control.

In Chapter 3, I formalise the refocusing principle and translate it into an optomechanical architecture. Much of the work presented here is in cooperation with Paul Meyer, who led optical design. My contribution is to derive a mathematical model of the tilted sensor and adjustable-focus air-lens; to source, select, and operate hardware components; and to design an enclosure that ensures portability and controlled imaging conditions. The resulting module is objective-agnostic, operates on standard collimated-beam camera ports, and requires no modification to the microscope's internal optical path.

In Chapter 4, I present a real-time control and acquisition software stack that closes the lens–sensor control loop at video rate. An algorithm is provided that determines the location of best focus (LoBF) on the tilted sensor in under 2 ms. This algorithm is robust to sparse edge distribution, brightness differences, and specular reflection. A C++ codebase achieves simultaneous operation of the cameras, LoBF algorithm, image registration, lens actuation, image rendering, image recording, and GUI interaction at 60 Hz or higher on commercial hardware. Finally, a software application with clinically useful and user-friendly features is presented, with a detailed manual in Appendix A.

Chapter 5 presents a comprehensive characterization of device performance. I measure closed-loop step responses, noting settling times to a range of realistic perturbations. I assess the device's impact on optical quality relative to a camera-only control, and confirm that the device achieves single-cell resolution. Finally, I evaluate the device's operating range across a variety of microscope platforms and objectives.

In Chapter 6, I apply the device to conjunctival imaging, and then demonstrate that it generalises beyond the slit-lamp. I acquire 30-second videos with two cameras simultaneously, one with and one without the autofocus device. This allows a direct comparison between videos with and without autofocus. I confirm single red blood cell (RBC) resolution and present an automated analysis pipeline to extract hemodynamic biomarkers, including vessel diameters, densities, and RBC velocity profiles. I then demonstrate portability to other microscopes, including a slit-lamp of a different model and a benchtop microscope, each with an infinity-corrected, collimated-beam camera port.

Chapter 7 concludes with a discussion of limitations, prospective improvements, and future directions. I outline opportunities in sensor fusion (e.g., combining the existing control loop with imaging-camera depth cues), stereo and multi-view extensions for improved depth recovery. Finally, I discuss pathways to wider clinical translation.

1.4 Publications and dissemination

To accelerate impact and invite early technical scrutiny, elements of this work have appeared in a patent, conference presentations, live device demonstrations, and internal symposia, as summarised below.

- **Patent.** Lynch, T.; Meyer, P.; Savin, T. *Automatic Autofocus System for Video Microscopy.* WO2025027199A1 (PCT), published 30 Jan 2025.
- **Manuscript (in preparation).** Draft article based on this thesis, currently being prepared for submission to *Nature Methods*.
- **BIOSTEC 2025 conference, Porto, Portugal.** 30-minute oral presentation: *"Video Autofocus Using an Adjustable-Focus Air-Lens and Sharpness Analysis of an Inclined Image Plane."* 20–22 Feb 2025.
- **BioMedEng24, London, UK.** Oral presentation (abstract track). 5–6 Sept 2024.
- **IP4U Universities Tech Fair, Imperial College London, UK.** Live device demonstration: *"Real-Time Autofocus for Video Microscopy."* 19–20 Sept 2023.
- **Internal Cambridge events.** Departmental research symposia and conferences (e.g., Internal Division C conference 2023 and 2024), with oral presentations and live demonstrations to clinical and engineering audiences.

Chapter 2

Literature review

2.1 Introduction

This chapter surveys literature relevant to an image-based, video-rate autofocus device for microscopy and its application to conjunctival imaging. For clarity and later use, we define a coordinate system where the optical axis is denoted as Z, with specimen or instrument motion in the X-Y plane orthogonal to it. These two motion classes have distinct implications for video microscopy. Motion in X-Y can be corrected with optical stabilization online [23] or image registration either online or offline [29, 76], at the modest cost of field-of-view loss at the image borders. In contrast, Z-axis drift alters the object distance and causes loss of high-spatial-frequency content that cannot, in general, be recovered from a single two-dimensional intensity image without restrictive assumptions about the point-spread function or sample structure [30, 121]. Accordingly, Z-axis motion must be mitigated during acquisition by an autofocus system.

Therefore, we adopt the following structure. Section 2.2 surveys autofocus strategies in microscopy, grouping them by how Z-axis focus error is sensed: *radiation-based* techniques infer object distance using auxiliary emitters and detectors, while *image-based* techniques infer defocus directly from the imaging beam. Section 2.3 reviews online and offline image registration techniques that address motion in the X-Y plane. Finally, section 2.4 synthesises the conjunctival imaging literature. Emphasis is placed on biomarkers that can be extracted from focused videos, associations with disease, as well as methodological limitations that motivate the need for video-rate, image-based autofocus.

2.2 Autofocus in microscopy

A central question at the outset of this project was whether to (i) develop the lab's novel tilted-sensor plus tunable-lens prototype, or (ii) pivot to an established commercial or published autofocus technique. This section surveys the design space with that decision in mind. Our aim is to both catalog approaches and evaluate their suitability for video-rate operation in constrained optical geometries (e.g. a slit-lamp), and for *universal* use at the camera port of microscopes without modifying their internal optical paths.

From a device-design perspective, desirable attributes include sufficient update rate for video-rate operation, low steady-state focus error, robustness to specimen artefacts, universality across a variety of imaging targets, and compatibility with camera-port integration. These criteria frame our discussion of autofocus methods, although not all metrics are available in the published literature.

Another way autofocus techniques are categorized in the literature is with the terms “active” or “passive” [65]. This thesis uses "radiation-based" and "image-based" instead, as these terms are more common in microscopy, and provide a clearer engineering distinction.

2.2.1 Radiation-based

Radiation-based autofocus infers Z focus error by sensing reflected, refracted, or scattered illumination incident on the imaging target. Commercial systems typically use near-infrared light projected through the microscope, which is reflected off a coverslip-sample interface [42, 102]. The measured return is converted to a focus-error signal that drives either a motorized stage or objective to hold object distance. As the sensing path is decoupled from the imaging beam, radiation-based approaches can run continuously at high update rates without interrupting acquisition.

Commercial radiation-based autofocus systems share a common architectural assumption: the presence of a planar reference interface near the specimen. Nikon's *Perfect Focus System* projects an 870 nm beam through the objective and records reflection off the coverslip surface using a detector [102, 42]. Typical implementations update the stage position up to 200 Hz and maintain sub-micrometre residual error [102, 42]. The radiation wavelength is chosen to avoid interference with common fluorophores and to minimize phototoxicity [102]. Zeiss' *Definite Focus* device follows a similar principle but with a projected 835 nm grid mask (rather than a beam) and detects reflection from the coverslip–sample interface rather than the coverslip surface [115]. The system explicitly relies on a refractive-index difference, so it is known to degrade or fail if mounting media cure to a glass-like index [115]. The Olympus *Z-Drift Compensation* device similarly relies on reflection from the

coverslip–sample interface, but uses a 790 nm diode beam and requires a glass-bottom dish of defined thickness (0.16–0.18 mm, 35 mm diameter) [60, 33]. As with all the aforementioned radiation-based autofocus modules, the *Z-Drift Compensation* device is microscope-specific, being mechanically integrated into the microscope frame [33]. In contrast, Applied Scientific Instrumentation's Continuous Reflective Interface Sample Placement *CRISP* device provides a portable radiation-based autofocus that mounts on a C-mount camera port. It employs a skewed illumination of the pupil with an IR mask in the range of 625–1050 nm wavelength, and detects reflections from either a glass–water or glass–air interface (from the manufacturer's manual). In other words, CRISP presupposes the presence of a reflective reference plane as well as a motorized stage or objective, as with all commercial radiation-based autofocus systems.

In academic literature, there are a variety of techniques available for radiation-based autofocus. One major motivation for the development of such techniques is to achieve nanometre-scale accuracy for super-resolution microscopy, as commercial systems are often insufficiently precise for this application [21]. Another motivation is to avoid expensive, microscope-specific components. To this end, Lightley et al. present an open-source radiation-based autofocus module called *openFrame* [66]. *openFrame* uses a dual-axis, cylindrical-lens optical autofocus device capable of single-shot, closed-loop operation with <200 nm accuracy. Similarly, Niederauer et al. present another open-source module, *qgFocus* ("quite good Focus"), which operates beyond the camera-port mount, similar to CRISP, but with off-the-shelf components and higher accuracy [85]. However, both these devices update the optical distance at low rates: 4 Hz for *openFrame* and 30 Hz for *qgFocus* [85, 66]. They also typically require high-numerical-aperture objectives, as they must be positioned close to the objective for sufficient laser stability [91]. To address the slow focus-adjustment rates, Rahmani et al. and Hsu et al. present astigmatism-based devices that introduce a small cylindrical lens into the probe beam, creating an intentionally astigmatic spot whose shape changes with defocus [38, 91]. When the sample is in focus, the spot is circular; when it moves above or below focus, the spot becomes elliptical along orthogonal axes. By measuring this ellipticity, the system can determine both the direction and magnitude of defocus. They achieve update rates of around 100 Hz in practice and operate across a wider range of 10–100× objective lenses. However, as with all commercial autofocus devices, the astigmatic approach relies on a consistent coverslip–sample interface. No literature was found on radiation-based autofocus devices used for slit-lamp microscopy.

In summary, radiation-based AF is a mature, precise, and fast solution when a planar reference is available and when the microscope can be engineered around the autofocus device. For video-rate microscopy in those conditions, the strengths of radiation-based

autofocus are clear. Defocus magnitude and direction is determined one-shot, update rates can exceed the frequencies of movement artifacts (>200Hz), and focus maintenance is robust to changes in image content as sensing does not rely on the imaging beam. These attributes explain why radiation-based systems are the standard solution for inverted microscopes performing high throughput or time-lapse imaging of adherent cells on glass [113, 111].

Yet those same systems rely on assumptions that do not generalize to most microscopes or imaging targets. First, these devices autofocus based on a reference coverslip surface, not the structure of interest. In multilayer, translucent, and/or moving specimens the AF reference can decouple from the imaging target, causing interface reflections to remain in-focus while the desired target drifts out-of-focus. Second, despite the use of near-infrared wavelengths, probe light can produce artefacts via reflections and scatter, and care is required to avoid contaminating sensitive detectors or specimens. Third, commercial modules exploit specific conjugate pupil or field planes deep inside the microscope body, where mechanical and optical integration is non-trivial.

These limitations become decisive in slit-lamp microscopy of the anterior eye and in conjunctival microvascular imaging. This environment involves multiple partially transparent layers and specular surfaces (tear film, cornea, conjunctiva, sclera), multiple layers of microvasculature, and involuntary motion in the Z axis. A coverslip-locking paradigm has no analogue. Engineering a bespoke, eye-safe system that discriminates between layers would require a complex system of emitters, detectors, and safety considerations, increasing cost and complexity without credible evidence that performance would be universal across subjects and targets. This motivates the next subsection on *image-based* approaches, which infer defocus directly from the imaging beam and can, in principle, lock to structures of interest without interacting with the specimen.

2.2.2 Image-based

Image-based autofocus estimates defocus directly from the imaging beam, with no range-finding radiation. Classical implementations with a single imaging camera will first modulate focus in a random direction, evaluate the change in image quality using a sharpness measure, and then step the optics toward an optimum [107, 24, 97, 119]. To improve the speed and universality of classical image-based autofocus, numerous sharpness measures have been proposed and compared in academic literature. See Section 2.2.3 for mathematical summaries of some high-performing measures.

There is also literature on improving the control algorithm itself, to minimize the time to find optimal sharpness. To that end, Yazdanfar et al. demonstrated a two-step method for digital microscopy that computes the slope of an image sharpness function from two frames,

and then steps to the estimated optimum [117]. Closely related developments in whole-slide imaging (WSI)—which acquires thousands of adjacent fields of view and mosaics them into a single composite image—illustrate that there are trade-offs between speed and accuracy at scale. Early WSI scanners pre-surveyed a focus map and then interpolated during scanning; modern WSI scanners have improved throughput using predictive strategies [78, 83]. Although substantially faster than exhaustive focus stacks, these and similar approaches still require a prior random modulation of optical power. As the sign of defocus is unknown *a priori*, these classical techniques necessarily exhibit focal-plane 'hunting'. This limits their effective application to video microscopy. In slit-lamp microscopy, this limitation is explicitly demonstrated by Rudolph et al. (2013), who applied state-of-the-art classical image-based autofocus to an electrically tunable lens and a camera attached to a slit-lamp [94]. They found that finding optimal focus took an average of 4 frames at 30 fps, so 133ms, despite using a stationary target [94].

Many groups have attempted to achieve one-shot image-based autofocus from a single imaging camera using deep learning. In computational photography, Herrmann et al. trained a convolutional neural network (CNN) to predict the correct lens position directly from one image, reducing mean absolute focus error by 3.6× compared to existing methods on handheld cameras [35]. In microscopy, Pinkard et al. (2019) and Liao et al. (2021) each applied a CNN to regress axial defocus from a single brightfield image, demonstrating accurate single-shot focus estimation [64, 90]. Zhang et al. (2024) also achieved single-shot focus estimation using deep reinforcement learning [120]. Data-driven strategies sidestep idiosyncrasies of hand-crafted sharpness metrics and can *estimate* both sign and magnitude of defocus one-shot. Their main limitation is the data requirement, which is in the thousands to hundreds of thousands of labeled frames for a single use-case. Generalization is also a major concern; even with large datasets, autofocus will typically fail on new objectives, unexpected illumination spectra, or different imaging targets [90, 64, 120]. Inference latency is also an issue, taking 50-59ms to determine optimal object distance from a single frame [90, 64]. These limitations currently preclude their use in video-rate microscopy, where rapid focus correction and cross-instrument generalization are essential.

A second family of image-based AF adds a secondary imaging plane of some form, thereby extracting the sign and magnitude of defocus from a single exposure [32]. These methods follow techniques used in photography, where dual-plane and dual-pixel phase-detection autofocus devices are commercially available in DSLRs and handheld cameras [17, 10]. In general, the technique in photography is to form two sub-aperture images and compute their relative lateral shift; the direction of the disparity gives the sign of defocus, and the magnitude of the disparity scales with object distance [17, 10]. An optical analogue

appropriate for microscopes is to split the pupil with a mask and form two images on a sensor; their relative shift provides defocus magnitude and direction. Silvestri et al. (2021) report a "rapid autofocusing via pupil-split image phase-detection" (RAPID) device, which uses a wedge plate to spatially separate the pupil. They demonstrate closed-loop focus correction on the order of a few hundred milliseconds for light-sheet microscopy, which illuminates samples with thin sheets of light to enable volumetric imaging [103]. Liao et al. (2016) developed a similar autofocus approach for WSI, using a two-pinhole aperture at the pupil plane to form two copies of the image. They then recover axial defocus from their translational separation [63]. Despite their effectiveness, both of these devices require access to a conjugate pupil plane for the split mask and sufficient optical real estate to relay two images. These have non-trivial integration requirements in compact or closed microscope bodies. In addition, RAPID corrects focus by translating the objective, and Liao et al. restore focus by translating the stage. While these actuation strategies are well-suited to light-sheet microscopy and WSI respectively, neither is practical for slit-lamp microscopy.

A final interesting approach to image-based autofocus is to capture multiple focal planes simultaneously, allowing refocusing post-acquisition [40]. Plenoptic and multi-focus light-field cameras (e.g., Raytrix designs) capture angular information sufficient to computationally refocus post-hoc. However, these systems impose substantial sensor and relay-optics constraints, and typically trade lateral resolution for sampling depth [40]. These constraints explain their limited adoption in mainstream microscopy. Jin et al. (2020) combine a wavefront modulating element with a deep neural network to perform extended depth-of-field microscopy, achieving a >5x increase in depth-of-field. However, the approach is trained to perform primarily with proflavine nuclear stained images. Reconstruction latency limits applicability to video operation, and data requirements mean universal camera-port integration is infeasible [47].

In summary, several strengths of image-based AF are clear. First, because sensing uses the imaging beam, the method inherently locks to image content rather than a surrogate interface. Second, image-based schemes successfully avoid phototoxicity and bleed-through concerns that arise in radiation-based AF. Third, where metrics are reported, state-of-the-art image-based systems achieve higher accuracy and larger working ranges compared to commercial radiation-based systems. Finally, single-frame strategies such as pupil-split phase detection provide the sign of defocus without lens hunting, enabling refocusing limited by camera frame rate and actuator dynamics [103, 63].

However, these strengths have not translated into a universal solution for autofocus, or for slit-lamp microscopy and conjunctival imaging. Classical image-based autofocus remains fundamentally unsuited to video-rate operation in the presence of continuous Z-motion,

even with attempts to apply deep learning. Single-frame pupil-split methods assume optical access to a conjugate pupil plane and space for additional relays [103]. Existing devices also actuate either the sample stage or objective, which doesn't generalize to slit-lamp microscopy. Learning-based approaches that are trained on narrow datasets struggle to generalize across objectives and imaging targets, and require large datasets and long inference times. No method is both universal and robust enough for the slit-lamp environment, which involves translucent multilayer tissue, specular surfaces, and involuntary Z-motion. These limitations motivate the design choices pursued in this thesis. We present an objective-agnostic, camera-port device that infers defocus sign and magnitude from the imaging beam at 60 Hz and above, and accommodates accordingly. The device attaches externally to the camera port without modifying the internal optical path, and is engineered specifically to tolerate brightness asymmetry and specularities encountered in conjunctival imaging.

2.2.3 Sharpness measures

As mentioned in the previous section, a core component of any image-based autofocus mechanism is a sharpness measure which can discriminate between in-focus and out-of-focus images. A number of existing papers in the field of microscopy attempt to identify top-performing sharpness measures with low computational costs [24, 28, 95, 77]. Unfortunately, the speed and accuracy of a sharpness measure partially depends on the type of image [77, 53]. In other words, no universally ideal sharpness measure exists, and empirical testing is required [77].

Therefore, four sharpness measures are identified in this section based on demonstrated performance across a variety of imaging targets. Each sharpness measure that is empirically tested in Section 4.3.2 is mathematically summarized here.

The Tenengrad measure is defined by Equations 2.1 and 2.2 below,

$$\begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I(i,j) = G_x(i,j) \qquad \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I(i,j) = G_y(i,j) \tag{2.1}$$

$$F_{\text{TEN}} = \sum_{i=1}^{M} \sum_{j=1}^{N} (G_x(i,j)^2 + G_y(i,j)^2) \tag{2.2}$$

where $*$ denotes convolution, $I(i,j)$ represents the brightness of pixel at position i, j in an image of dimensions (M,N), and F_{TEN} represents a scalar sharpness score [100]. This measure was identified by Shah et al. as a top-performing function for imaging sputum smears [96]. It works by convoluting the image pixel array against two $3x3$ Sobel matrices

(Eq. 2.1) to detect sharp intensity changes in the horizontal and vertical direction. Then, the resulting two matrices are squared and summed to get a final score (Eq. 2.2).

The Vollath F_4 measure is defined by Equation 2.3,

$$F_{\mathrm{VOL}} = \sum_{i=1}^{M-1} \sum_{j=1}^{N} I(i,j) \cdot I(i+1,j) - \sum_{i=1}^{M-2} \sum_{j=1}^{N} I(i,j) \cdot I(i+2,j) \tag{2.3}$$

where F_{VOL} represents a scalar sharpness score [77, 95]. This measure was identified by Mateos-Pérez et al. and Santos et al. as a top performing function for imaging tuberculosis bacteria and stained erythrocytes, respectively [77, 95]. It works by computing the image autocorrelation, which has maximum magnitude at high intensity gradients. Vollath F_4 is also said to be robust against noise-induced intensity variations [77, 95].

The Roberts Cross sharpness measure is defined by Equation 2.4 [44],

$$\begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} * I(i,j) = G_x(i,j) \qquad \begin{bmatrix} 0 & +1 \\ -1 & 0 \end{bmatrix} * I(i,j) = G_y(i,j) \tag{2.4}$$

and then Equation 2.2. This measure was identified by Jhapate et al. as high-performing for retinal microvascular imaging, where features can be particularly thin [44]. It is computed the same way as the Tenengrad measure, but using $2x2$ Roberts Cross operators instead of $3x3$ Sobel matrices. The Roberts Cross measure is computationally faster, but is more susceptible to noise [44].

The Canny sharpness measure is defined by Equations 2.1 and 2.2 [9] when used as a sharpness measure rather than an edge detector. Canny works the same way as Tenengrad, but an upper and lower threshold is applied to the result of Equation 2.1 [9]. Unlike the other measures, a Canny function is provided with the OpenCV C++ library.

2.3 Image registration

Movement artifacts in the X-Y plane can be managed with optical registration during acquisition, or with image registration during or post-acquisition. While the main focus of this thesis is on accommodating movement artifacts along the Z axis, Section 2.3 presents an online image registration pipeline that accurately stabilizes 1944 × 1472 resolution images at video rate. This pipeline works up to 120 Hz in parallel with the autofocus functions (analysis of the tilted sensor, control of the lens actuator) and GUI processes (rendering, recording, and playback). This enables 3D X-Y-Z stabilization on commercial hardware, which is a non-trivial computational challenge. We also present a post-acquisition pipeline for

processing conjunctival imaging videos which involves offline image registration. Therefore, in this section we briefly survey state-of-the-art registration techniques.

As noted in Section 1.2, Tom Drummond's early software for the Prosilica GC1380H camera demonstrated real-time stabilization of 1360 x 1024 images at 30 Hz on legacy hardware. This was done by detecting vessel edges and tracking their motion with one-dimensional searches perpendicular to the vessel axis [92]. This work anticipated developments in feature-based image registration, where matching local edges or feature descriptors across frames is used to estimate global motion. This is one of three families that dominate the field: intensity-based matching, feature-based matching (as described by Drummond), and frequency-based matching [125].

Intensity-based matching such as the classical Lucas–Kanade (LK) alignment minimizes a photometric error under a parametric warp [4]. Updates to LK achieve sub-pixel accuracy with a Gauss–Newton step; multiscale (pyramidal) variants of this algorithm are now a workhorse for real-time stabilization [71, 4, 110]. However, LK is fragile to changes in brightness or asymmetries in illumination. Instead, the Enhanced Correlation Coefficient (ECC) formulation maximizes a normalised correlation that is invariant to linear brightness changes and converges reliably under similarity/affine models [22]. These properties make ECC attractive for slit-lamp videos with illumination asymmetry, although the technique is computationally expensive [22]. Gradient-domain alignment (minimising differences of spatial derivatives) is another practical variant in vessel-rich scenes, as it suppresses low-frequency illumination bias [109].

Feature-based pipelines detect salient points or curves in the scene, describe them, match across frames, and estimate a global transform. In microscopy, algorithms based on corners (Harris/Shi–Tomasi) and multi-scale blob features (SIFT/SURF/ORB) are common [99, 70, 5, 93], followed by global motion estimating using RANSAC to remove outliers [25]. In conjunctival imaging, vesselness pre-filters (e.g., Frangi) can emphasize curvilinear structures before feature matching or before Drummond-style 1-D edge tracking along normals to centre-lines [27, 43]. Feature-based pipelines are robust to local intensity fluctuations (moving RBCs) and occlusions (blinks). However, they often require user-input to calibrate feature-detection sensitivity, and they fail when images have texture-poor regions.

Frequency-domain matching essentially refers to phase correlation following Fourier analysis. For near-pure translations, frequency-domain phase correlation localizes the cross-power spectrum peak to yield displacement. It is intrinsically insensitive to uniform illumination changes and, even with sub-pixel refinements, is both fast and accurate [1, 31]. In practice it is often combined with ECC or LK to include small rotations/scale [1, 31]. This method is increasingly common due to it's effectiveness across imaging environments,

While matching frames is the first step, one must also select a geometric model to optimize for. In video-rate microscopy, the most common geometric models are pure translation, rigid (rotation+translation), similarity (rigid+isotropic scale), affine (adds shear and anisotropic scaling), and planar projective (homography). In-practice, online image registration necessarily emphasizes computational efficiency, so rigid or pure translation models are preferred. Offline, registration accuracy takes precedence over computational speed so affine and homography models are preferred. This is the case in conjunctival imaging; rigid models are typically sufficient to compensate microscope tremor, microsaccades, and small out-of-plane tilts over short time horizons. However, cautious non-rigid refinement of registration is necessary for effective quantification of conjunctival videos, as the translucent ocular surface often rotates and shears over time. Feature-based and intensity-based matching have been applied to the full range of geometric models in the literature, while frequency-based matching works best with rigid registration.

In conjunctival-specific literature, a notable contribution is provided by Felder et al. (2016) who implemented *optical* stabilization on a slit-lamp microscope, thereby avoiding data loss due to field-of-view truncation that occurs with all image registration [23]. They achieved stabilization up to 90 Hz using cross-correlation between frames, an intensity-based matching algorithm. Khansari et al. combine offline intensity-based matching with frame exclusion heuristics to extract stable sequences from larger videos of the conjunctiva, using a similarity (rigid + isotropic scale) geometric model [50]. Brennan et al. use a similar pipeline, first extracting in-focus segments from 5-15s videos and then using an offline intensity-based image matching technique with an affine geometric model [8, 7, 26]. Exemplar in-focus segments were just 0.5-1s long [8]. Yun et al. (2022) provide a post-processing pipeline for conjunctival videos, using offline intensity-based matching with ECC and a pure translation geometric model to extract bloodflow velocity profiles [118]. In all cases, Z axis movement artifacts were a clear limitation, as all groups necessarily extracted small sections from longer videos to perform quantitative analysis of varying effectiveness.

Taken together, the literature demonstrates that both online and offline image registration are valuable for stabilising motion in the X–Y plane. Online registration methods—typically being phase correlation or feature-based tracking—can run at video-rate and support autofocus in dynamic environments. Offline methods, though slower, allow for higher-order corrections that improve quantitative analysis. Neither class of method, however, can correct axial (Z-axis) drift, which irreversibly degrades high-frequency image content. This distinction reinforces the need for an autofocus device that can be integrated with registration pipelines to achieve full three-dimensional stabilization of conjunctival video microscopy.

2.4 Conjunctival imaging

The microvasculature comprises the most distal and numerous vessels of the vascular system, perfusing almost every living tissue in the human body. Crucially, microvessels of diameter ~8-100 μm serve as the body's blood-tissue interface, allowing tissue oxygenation, nutrient exchange, reparative processes, and hormone signalling [112]. Healthy microvasculature is organized hierarchically to form circulatory circuits, known collectively as the microcirculation. Microcirculatory dysfunction is associated with many diseases, including diabetes [51, 13, 3, 62, 16, 69, 14], hypertension [105, 84], cardiovascular disease [2, 8, 55], glaucoma [72, 52, 108], and thyroid syndrome [81], which are considered further in Section 2.4.3.

The anterior segment of the human eye has historically attracted researchers due to the relative ease at which numerous microvascular networks can be imaged non-invasively. It is one of the few locations where the microcirculation may be observed with the naked eye, and uniquely presents complex two-dimensional networks of microvessels embedded within translucent tissue. Indeed, the first published studies on the human microcirculation were conducted by Herman Boerhaave (1668 - 1738), and he used the conjunctival microvasculature [112].

In ocular research, it is common to assume that observations from conjunctival microcirculations are applicable to the rest of the body. This assumption has been tested in porcine [58, 59], ovine [36], and rat models [122], and has been found to be largely true. A 2020 study observed that changes to conjunctival microcirculations in hypothermic rats were closely related to vascular changes in the cerebral cortex [122]. Based on this assumption, research involving conjunctival imaging is responsible for much of what we know today about the microcirculation throughout the human body. Our contemporary understanding of capillary angiogenesis mediation, for example, is largely based on observations from the conjunctival microvasculature [79, 75].

Another assumption which is fundamental in conjunctival imaging research is that the measurements are repeatable and sufficiently unaltered by external factors. Xu et al. (2015) found that there was no significant change in blood flow velocity, blood flow rate, and mean vessel diameter for the same conjunctival microcirculations imaged at 9am and 5pm [116]. Critically, conjunctival vessel morphology is also sufficiently static on the scale of weeks to months [101]. These findings mean that clinicians and researchers can return to specific vascular structures following a course of treatment, and can attribute microcirculatory changes to treatment outcomes.

2.4.1 Techniques

Techniques used to image conjunctival microvasculature in published literature include Hemoglobin Video Imaging (HVI) [72, 81, 52, 84, 80], fluorescein angiography [82], high speed microcinematography [57], handheld microscopy [36, 122], handheld photography [62, 3], functional slit-lamp biomicroscopy (FSLB)[101, 46, 116, 69, 98, 118], diffuse reflectance spectroscopy [58, 59], orthogonal polarization spectral imaging (OPSI) [114], computer-assisted intravital microscopy [14, 15, 13, 104], and even retinal functional imaging applied to conjunctival microcirculations [45]. Many of these names refer to essentially the same methodology due to the lack of standardization across fields where the conjunctival microcirculation is imaged. For example, HVI, FSLB, and OPSI all use predominantly green-wavelength light paired with slit-lamp microscopy [101, 114]. Some researchers have even given commercial names to their conjunctiva imaging systems, such as the Eyeflow device from the University of Illinois [50, 51] and the ODIN concept from ODI Medical, Norway [59, 58].

All of these techniques essentially produce high-contrast images of conjunctival microvasculature with varying degrees of cost, resolution, and invasiveness. Specific microvasculature structures can be revisited in each patient, which enables clinicians and researchers to observe disease evolution and/or the effects of various treatment options. No technique produces images with higher resolution or contrast than HVI, which is the term used throughout this report. HVI can be performed at a lower cost than all but handheld microscopy and photography, and is less invasive than fluorescein angiography.

The HVI technique was first reported by Paul A. R. Meyer in 1989 [79]. The name refers to a band-pass filter which transmits primarily 505-575nm (green) wavelength light, which is added to the slit-lamp's factory illumination system [79]. This wavelength corresponds to one of the absorption peaks for haemoglobin [79]. A hot mirror is also added, which blocks wavelengths above 730nm from reaching the imaging camera [79]. These minor modifications to the slit-lamp allow high-contrast, high-resolution videos to be recorded, which show dark erythrocytes against a light, reflective background of sclera. A sample image produced using HVI is shown in Figure 1.1.

All of these imaging techniques are challenged by constant involuntary movements of the patient's eye. With conjunctival imaging, movements in the X-Y plane are largely due to microsaccades and patient movement. Z axis movements are typically due to muscle action, respiration, and heart beat, and they cause constant defocusing of the imaged microcirculations. Due to these movement artifacts, none of the techniques described above can reliably and repeatedly produce in-focus videos longer than two seconds. Therefore, analysis of

conjunctival microcirculations in existing literature necessarily focuses on images or very brief videos.

2.4.2 Biomarkers

The quantitative measures that researchers frequently extract from imaged conjunctival microcirculations are vessel diameter (D), axial flow velocity (V_a), and capillary density. From these, cross-sectional (or, average) flow velocity (V_c), vessel flow rate (Q), wall shear rate (WSR), and wall shear stress (WSS) can be derived using well-established formulae [8, 86].

D is generally taken at the full-width at half-maximum point of an intensity profile perpendicular to the vessel. A similar technique can be used to determine the diameter of an aqueous column inside a venule, which is shown in Figure 2.2.

V_a is almost always determined by creating a space-time image (STI) that shows the change in pixel intensities along an identified vessel over time [86]. An example of an STI is shown in Figure 2.1, where moving erythrocyte aggregates cause linear streaks with a slope of V_a [69]. V_a usually varies in microvessels due to cardiac impulse, intra-ocular pressure change, and other phenomena. If a vessel has periodic changes in V_a, this is known as the velocity pulse period (VPP) [57]. The accuracy of a V_a measurement depends strongly on maintaining vessel focus for about two to four seconds [57]. This requirement makes research relying on V_a cumbersome and inaccurate [86], requiring multiple trials and a trained slit lamp operator. Even then, V_a is only successfully extracted from vessels of a particular length and diameter [46].

V_c is usually derived from V_a using equations based on empirical data from representative models of microvasculature. An example of one equation is below, which is based on the ratio of the vessel diameter (D) to the erythrocyte diameter (D_c) [56]:

$$V_c = \begin{cases} V_a & \text{when} \quad D/D_c \leq 0.6 \\ \frac{V_a}{1.58(1-e^{-\sqrt{2D/D_c}})} & \text{when} \quad D/D_c > 0.6 \end{cases}$$

Then, vessel flow rate (Q), wall shear rate (WSR), and wall shear stress (WSS) can be derived from D and V_c using established formulae [8]:

$$Q = V_c \frac{\pi D^2}{4}$$

$$\mathrm{WSR} = 8\frac{\mathrm{V_c}}{\mathrm{D}}$$

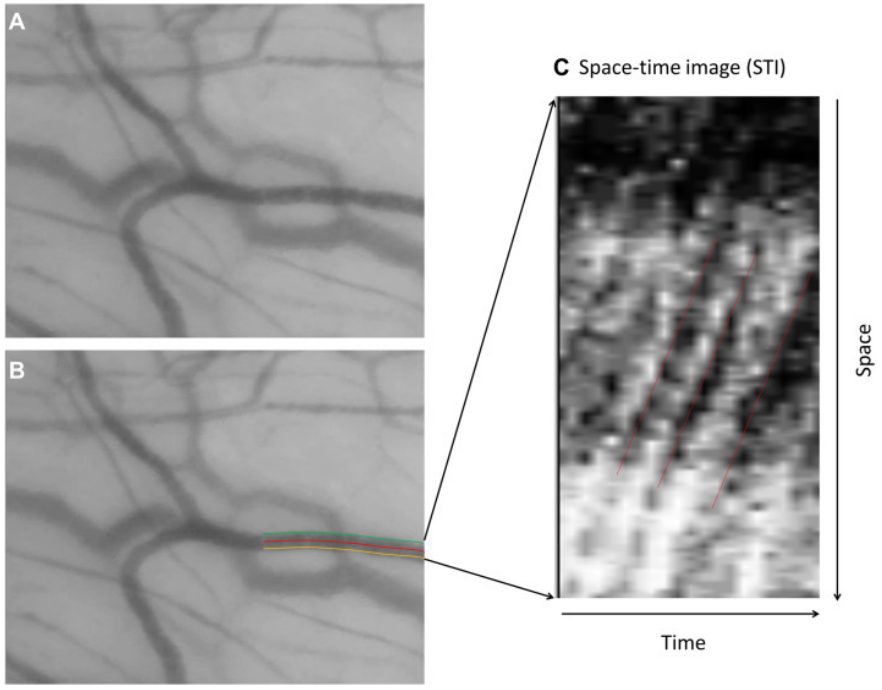


Fig. 2.1 Example of a space-time image (STI) used to calculate V_a from a conjunctival angiogram, from [69].

$$\text{WSS} = \mu \times \text{WSR}$$

where μ is the dynamic viscosity of blood, which is usually a function of D.

Capillary density isn't quantified in a standardised way. Density can refer to the number of capillaries which cross an arbitrary straight line [58], or the length of capillaries in a $1mm^2$ area [105]. An automatic approach to determining capillary density will use Frangi filtering, adaptive thresholding, or U-net segmentation to identify capillaries [118]. Then, a density score with arbitrary units is usually provided.

2.4.3 Clinical motivations

A focused review of existing clinical applications of conjunctival imaging is necessary to judge the utility of the technique and to quantify how autofocus could extend its diagnostic reach. Theoretically, conjunctival imaging is applicable to any disease or pharmacological intervention that affects the microvascular system. In practice, using literature published after 2005, conjunctival imaging has been used to research diabetes (and especially diabetic retinopathy) [3, 62, 51, 50, 69, 13, 15], Alzheimer's disease [104], sickle cell disease [14], dry eye disease (DED) [12, 101, 11], contact lens (CL) wear [101, 39, 46], various cardiopul-

monary bypass techniques [58], cardiovascular disease (CVD, in various forms) [8, 55, 2], glaucoma [108, 48, 72, 52], hypertension [84, 105], and thyroid eye disease [81].

Not every clinical application is covered in detail in this literature review. Some diseases and procedures have only been approached through preliminary pilot studies, which is true for Alzheimer's [104], cardiopulmonary bypass [58], and sickle-cell disease [14]. Additionally, conjunctival imaging has only taken a partial role in larger investigations of thyroid eye disease [81] and hypertension [84, 105]. In fields with only one or two papers involving conjunctival imaging, it is difficult to determine the value that HVI and autofocus can provide. Therefore, these diseases and procedures are not discussed at length here.

Instead, CVD, DED, CL, glaucoma, and diabetes are considered in detail in this section. Compared to other applications, each of these fields has a large volume of research involving conjunctival imaging. Each disease is briefly described, and findings related to the conjunctival microcirculation are summarised. In each section, papers are highlighted if they indicate a need for longer videos of conjunctival bloodflow, as this demonstrates a need for autofocus. Studies are also emphasised if they suggest that HVI has a strong chance of discriminating between healthy and ill patients, as this demonstrates clinical value.

Cardiovascular Disease (CVD)

CVD accounts for 17.9 million deaths globally every year, and is the leading or second leading cause of death in all developed countries [2]. Research applying conjunctival imaging to CVD tends to focus on coronary artery disease (CAD), the main type of CVD [55], and on patients who have recently suffered from acute myocardial infarction (MI) [8, 2], one of the major symptoms of CVD. However, cyanotic congenital heart disease (CCHD) in adult patients has also been examined [8]. Generally, researchers find that patients suffering from any form of CVD have a significant but minor reduction in V_a and WSR compared to control groups, although error bars are large for measurements involving V_a [8, 2]. Changes in D are either insignificant or slightly increased, and changes in Q are either insignificant or slightly reduced. [8, 2].

A common research aim in this field is to use conjunctival imaging to screen patients for vulnerability to coronary sclerosis [55]. For example, Awuah et al. published a case–control study in April 2022 comprising 132 patients, using smartphone-assisted slit-lamp videography of the conjunctival microcirculation to measure haemodynamic parameters. They proposed a CVD screening algorithm combining V_a measured using a smartphone with biomarker measurements from a blood sample [2]. This supervised machine learning algorithm achieved an area under the receiver operator characteristic curve (AUROC) of 96.7% [2], indicating that the algorithm correctly ranked post-MI patients above healthy controls about 97 percent

of the time. For comparison, typical AUROC scores for a COVID-19 rapid test compared to PCR test are 89.1% to 96.8% [89].

One flaw with Awuah et al's algorithm was that it tended to generate false-positives due to confounding factors in the control group [2]. Another problem was that patient's V_a changed significantly throughout a cardiac cycle, and the researchers couldn't reliably maintain focus on patient's conjunctival vessels for longer than one cardiac cycle. HVI paired with autofocus would solve this problem, and would therefore be valuable in CVD research. However, existing evidence suggests that it isn't viable to develop a clinically useful screening tool based solely on conjunctival imaging. HVI data would need to be combined with other metrics, such as blood-borne biomarkers or electrocardiogram readings, to successfully screen for CVD.

Dry Eye Disease (DED) and Contact Lens (CL) Wear

Estimates of DED prevalence range from 5 to 34 percent of the population [12]. The disease is characterised by a loss of homeostasis of the tear film, causing minor to severe ocular discomfort and pain. DED is usually associated with CL wear, as they are co-pathologic, although DED can also be caused by immune dysfunction. DED is currently diagnosed based on the patient's reported symptoms, as there is no gold-standard diagnostic test [11].

The conjunctival haemorheologic outcomes for DED and CL wear are very similar, so they are reported together here. All conjunctival research on acute CL wear, systemic CL wear, and DED has found significant increases in capillary density, D, V_a, and Q. Shu et al. postulated that both systemic CL wear and DED cause chronic ocular inflammation, which leads to the observed changes in conjunctival microvasculature [101]. Similarly, acute CL wear is thought to cause acute inflammation and loss of blood perfusion to the limbus, which causes increased conjunctival blood flow as an adaptation [101]. HVI could theoretically be used to improve CL design by ensuring minimal impact on the conjunctiva, limbus, and cornea.

Chen et al. published a study in July 2018 comprising 50 patients, which used FSLB and an algorithm based on Q, D, V_a, and capillary density to screen for DED [12]. This supervised machine learning algorithm achieved an AUROC of 90.6% when compared against the ocular surface disease index (OSDI) diagnostic tool [12]. Interestingly, the OSDI tool itself only achieves AUROC scores of 62% to 96% when compared against other diagnostic tools [12], meaning that conjunctival imaging has the potential to become a diagnostic criterion. However, whether it will become a clinically useful criterion partially depends on clinical need. The study authors also noted that the algorithm would improve significantly if blood biomarkers such as inflammatory cytokines were collected in addition

to conjunctival haemorheologic measurements [12]. Ultimately, it seems unlikely that HVI could lead to a clinically useful diagnostic criterion for DED. Instead, conjunctival imaging techniques like HVI will continue to be valuable research tools for DED and CL wear research.

Glaucoma

Glaucoma is the leading cause of irreversible blindness, with an estimated 3.6 million people blinded globally due to glaucoma in 2020 [106]. The primary cause of glaucoma is chronically high intra-ocular pressure due to blockage or excess resistance in the aqueous drainage system [106]. Aqueous outflow occurs primarily through the trabecular meshwork into Schlemm's canal, and then into limbal episcleral veins [48, 49]. These episcleral veins are known as aqueous veins, because the higher-velocity aqueous forms a centralised column within a lower-velocity sheath of blood flow. This is shown in Figure 2.2A from Lusthaus et. al, below [72]. When aqueous veins are imaged using HVI, they appear as a centralised erythrocyte void, as shown in Figure 2.2B [72]. The diameter of this aqueous column can be determined using the pixel intensity distribution perpendicular to the aqueous vein, shown in Figure 2.2C [72]. This is similar to how D is determined for any capillary in an image of conjunctival microvasculature.

The application of HVI to glaucoma research is in its infancy. No existing research has collected and compared haemorheologic parameters like V_a and Q for patients with and without glaucoma. No supervised machine learning algorithm for glaucoma diagnosis has been attempted, as it has for CVD [2], DED [12], and diabetes [50, 62, 3]. Although glaucoma doesn't affect blood rheology directly, it may be valuable to apply techniques from other fields to HVI videos of patients with glaucoma.

Instead, existing HVI-glaucoma research focuses on the aqueous column. Typically, the main quantitative measure is column diameter, determined using the technique from Figure 2.2C [72]. Based on this measurement and qualitative observations, aqueous outflow was shown to vary with changes in intra-ocular pressure [52, 48]. The velocity of an aqueous column hasn't been measured in existing literature using HVI [52]. This measurement should be possible, either by measuring the length of an aqueous column, or by analysing velocity changes in surrounding erythrocyte aggregates using an STI (as shown in Figure 2.1). To be accurate, such a measurement would require high-resolution, in-focus videos lasting four seconds, since aqueous velocity is pulsatile [52, 48]. HVI paired with autofocus can uniquely and non-invasively provide such videos.

A common treatment for slowing the progression of glaucoma is a trabecular bypass, which is where a small incision is made in the sclera to relieve intra-ocular pressure and

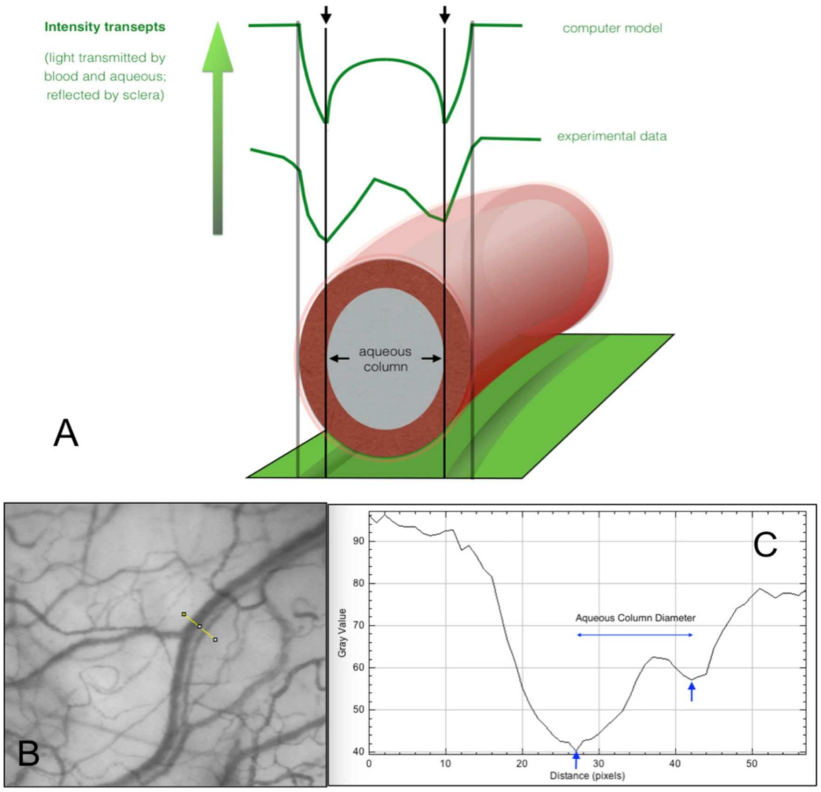


Fig. 2.2 (A) A modeled aqueous vein; (B) an HVI image showing an aqueous vein; (C) a pixel intensity graph taken perpendicular to the aqueous vein [72].

allow aqueous drainage [49]. Lusthaus et al. applied the HVI technique to partially quantify outcomes of such a procedure [73]. HVI successfully demonstrated increased aqueous outflow post-operation in this study [73]. Lusthaus et al. also outlined the need for further noninvasive assessments of aqueous outflow *in vivo*, which can be provided using the HVI technique [73]. An improved HVI system would be ideal for evaluating clinical treatment options for patients with glaucoma.

Finally, it seems unlikely that HVI could become a clinical screening or diagnostic tool for glaucoma the way it could for diabetes or CVD. Since glaucoma can arise in subjects with normal intra-ocular pressure [49, 48], confirming normal aqueous outflow isn't a guarantee against glaucoma development. Instead, a mature HVI system would be applied to glaucoma research, and to confirming the effectiveness of treatment.

Diabetes

Diabetes mellitus affects an estimated 451 million people globally according to an estimate in 2017, which is projected to increase to 693 million people by 2045 [16]. It secondarily affects

the peripheral vascular system, and is characterised by poor blood sugar regulation due to a lack of insulin (type 1) or a resistance to insulin (type 2) [16]. Diabetes complications affect the kidneys, feet, and eyes, potentially causing MI or stroke [16]. In fact, many people suffering from diabetes are screened annually for diabetic retinopathy using retinal function imaging [62]. Perhaps for this reason, there is a large volume of literature where conjunctival imaging is applied to diabetes research. Conjunctival imaging doesn't require retinal dilation and is more accessible than the retinal fundus.

Generally accepted haemorheologic markers of diabetic conjunctival microcirculations compared to healthy controls include significantly reduced V_a, lower WSS, lower WSR, and reduced capillary density [50, 86]. An additional unique marker of diabetic conjunctival microcirculations is a significant increase in the prevalence of microvascular abnormalities [15, 13, 50]. Diabetes researchers have named specific abnormalities and labelled images with a severity index based on the prevalence of these abnormalities [13]. Figure 2.3 is reproduced from Cheung et al. as an example of some named abnormalities that are correlated with type 1 diabetes [13].

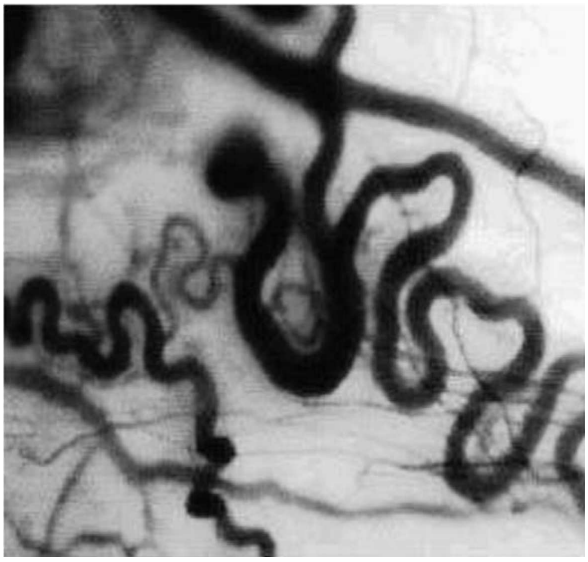


Fig. 2.3 Image of conjunctival microvasculature with abnormally large diameters, excessive tortuosity, and a ruptured microaneurysm, all markers of diabetic disease [13].

Based on the identification of abnormal morphology, several researchers have applied supervised machine learning and deep learning to discriminate healthy and diabetic conjunctival microcirculations [3, 62, 50]. In a 2016 study with 76 patients, Khansari et al. used slit-lamp microscopy, fine-structure image analysis, and a supervised machine learning algorithm to achieve AUROC scores of 72% to 97% depending on the stage of diabetic

retinopathy [50]. Notably, two expert human retinal specialists achieved AUROC scores of between 54% to 67% with the same images.

Building on that study are two papers published in January and March of 2022, by Li et al. and Babenko et al., respectively [62, 3]. Both trained deep learning models to discriminate stages of diabetic retinopathy from low magnification photos of the whole eye. The January study used 611 images from 180 participants as a training dataset [62]. They achieved an average AUROC score of 82.0%, and the algorithm's accuracy was significantly higher compared to four trained ophthalmologists (75.2% compared to 51.6%) [62]. The March study from a research team at Google Health used photos from 145,832 patients as a training dataset [3]. They achieved an AUROC score of between 75.0% and 79.2% in determining the clinical stage of diabetic retinopathy [3]. For comparison, logistic regression models based on sex, ethnicity, and years with diabetes achieved AUROC scores of 64.8% to 66.5% [3].

These supervised machine learning and deep learning algorithms achieved better-than-human predictive capabilities with low-magnification, low-contrast photographs. They were also entirely based on microvascular morphology, with no way to determine useful parameters like V_a, WSS, and WSR. A training database of high-quality focused HVI videos would provide better data on both microvascular morphology and haemorheologic parameters. An algorithm trained on such data could potentially achieve equal or better performance. This is discussed further in Section 7.3.

2.5 Conclusions

Taken together, the literature makes the central problem plain: video-rate microscopy fails when axial motion is left to post-processing. Registration in the image plane, whether intensity-, feature-, or frequency-based, stabilises X–Y effectively but cannot recover high-frequency content lost to Z-defocus. Radiation-based autofocus excels when integrated into a bench microscope that assumes a coverslip reference, but it performs poorly when generalised to other microscopes and specimens where a coverslip reference isn't available. It certainly is inapplicable to the multilayer, translucent, specular, and coverglass-free environment of a slit-lamp. Classical image-based autofocus, meanwhile, still "hunts" because the sign of defocus is unknown *a priori*, and learning-based single-frame estimators remain data-hungry, slow, and brittle to domain shift. Pupil-split and dual-plane image-based autofocus techniques show promise for one-shot refocusing, but remain slow and application-specific. Against this backdrop, the case is compelling for a one-shot, image-based autofocus device that

locks to the actual imaging content, runs at video rate, and requires no internal microscope modifications.

The clinical and research implications motivate this work. Reliable, in-focus videos unlock robust measurement of capillary density, vessel diameter, RBC velocities, wall shear rate, wall stress, and flow rate. It allows measurement of such biomarkers over whole cardiac and respiratory cycles, rather than from cherry-picked sub-second segments. That capability directly addresses pain points in research into cardiovascular disease, dry eye disease, glaucoma, and diabetes. In glaucoma in particular, continuous focus would allow non-invasive quantification of aqueous column dynamics. In diabetes and cardiovascular disease, it would improve the measurement of blood flow pulsatility and microvascular abnormality metrics. More broadly, such a device would provide a practical means for longitudinal, non-invasive monitoring of the microcirculation in response to disease and treatment.

For these reasons, this thesis proceeds with a tilted-sensor plus tunable air-lens architecture. This chapter's review also clarifies what such a device must achieve. It must determine both sign and magnitude of defocus from a single exposure, update at or above camera frame rate (60 Hz or higher), and be amenable for routine use by non-specialist clinicians. In addition, it must remain effective despite illumination asymmetry, sparse texture, and specular reflection, and integrate universally at a microscope's camera port without disturbing the optical train. To achieve the latter, the device must refocus by actuating optical power rather than the stage or objective, for contexts where neither can be moved. Algorithmically, the sharpness measures canvassed here provide tools for fast, edge-centric estimation of spatial sharpness across the tilted sensor. The comprehensive consideration of image registration techniques suggest computationally-efficient methods for online real-time registration, and highly accurate methods for offline registration post-hoc. In sum, the literature indicates the feasibility of a device that maintains focus along the Z-axis while stabilising motion in X and Y, enabling three-dimensional video-rate stabilization on commercial hardware.

The next chapter translates the requirements distilled in this chapter to a concrete system architecture. We formalise the refocusing principle, select and place components in the collimated beam, and design the enclosure and interfaces. The outcome is a fast, accurate, and objective-agnostic module which generalises to other devices and is practical for clinicians to operate.

Chapter 3

System architecture and component selection

3.1 Introduction

This chapter translates the requirements distilled in Chapter 2 into a concrete system architecture for a universal, video-rate, image-based autofocus module. The autofocus mechanism is based on a tilted sensor for depth sensing paired with a tunable air-lens for refocusing, as originally conceived by Paul Meyer. Section 3.2 explains this mechanism in detail alongside an optical ray diagram. The same section develops mathematical models for the tunable air-lens and the depth-sensing camera, and validates the models with empirical measurements.

Sections 3.4–3.6 then discuss component-wise design decisions for the crucial elements: the tilted sensor, refocusing actuator, imaging camera, and computing platform. Components are evaluated against alternatives, with bench tests that compare feasible options against the three design goals introduced below. Section 3.7 describes four iterations of the light-tight enclosure, necessary for robust operation in brightly lit clinical environments, and briefly details the adapter required to transfer the device from the slit-lamp to a benchtop microscope, supporting the generalization discussed in Chapter 6. In sum, each hardware decision that constitutes the system architecture is justified before turning to the real-time acquisition and control software in Chapter 4.

Three goals guide all design choices and component selections. First, *speed:* the control loop must operate at or above video rate (60 Hz or higher), with sensing, computation, and actuation budgets that sustain closed-loop operation during typical ocular motion. Second, *accuracy and robustness:* it must reliably converge across objectives and conditions, tolerate artefacts such as brightness asymmetry or specular reflection, and remain stable under

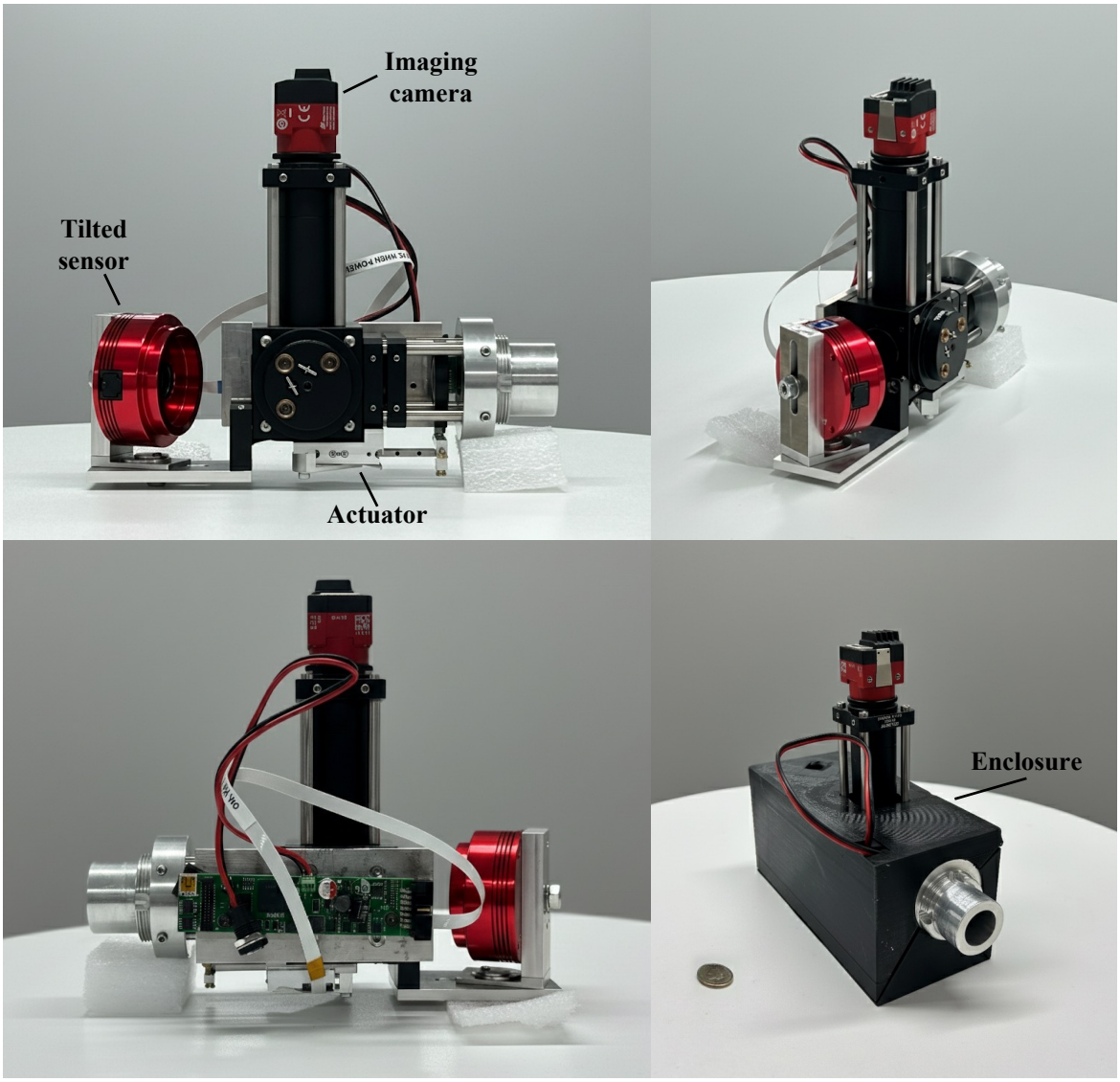


Fig. 3.1 Real photos of the autofocus device discussed in this thesis. A front view (top-left), oblique view (top-right), back view (bottom left), and another oblique view with the light-proof enclosure and a pound coin for scale (bottom right).

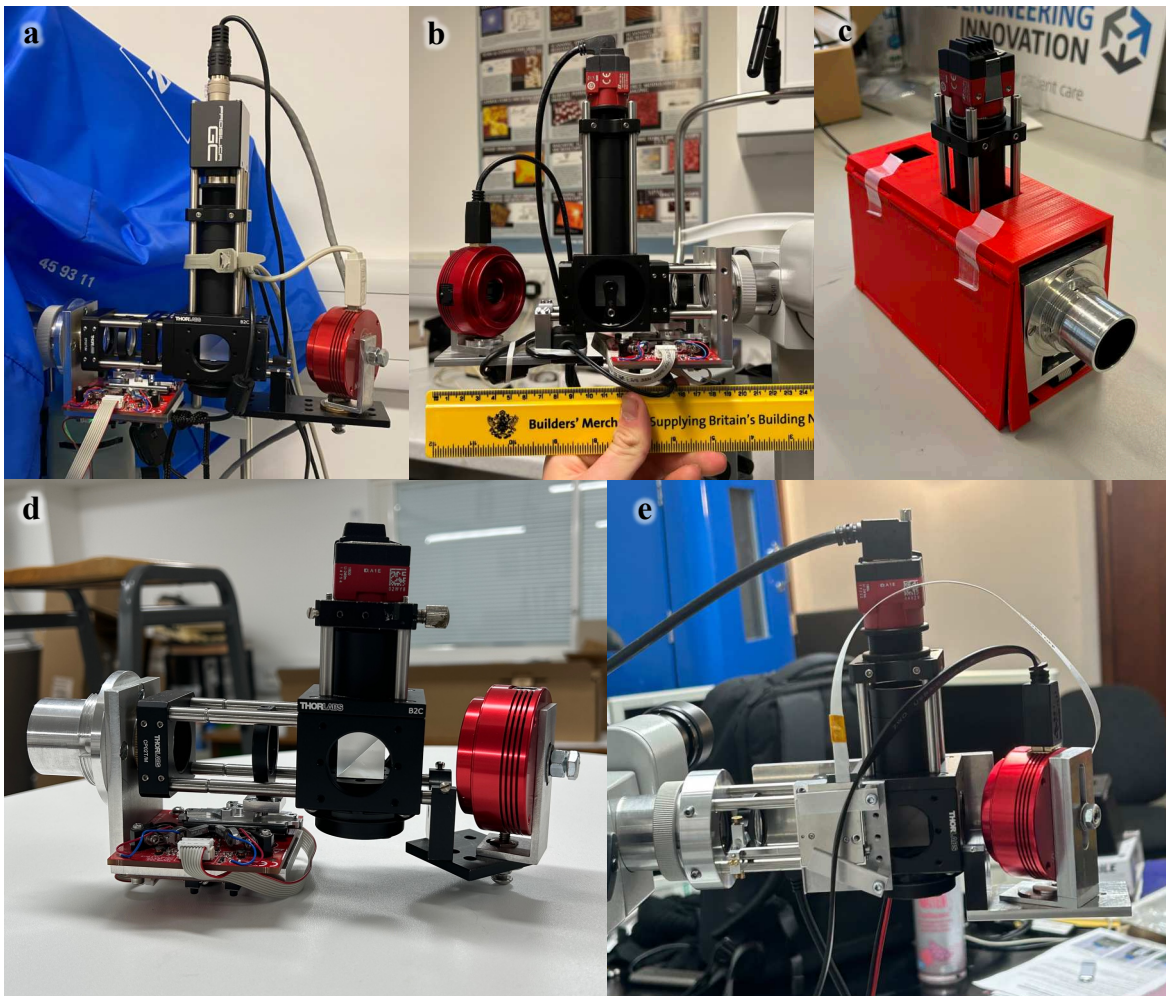


Fig. 3.2 Photos of earlier prototypes: (a) earliest photo of the device; (b) iteration with a new imaging and tilted camera; (c) the working prototype in the Clifford Albutt lab; (d) an attempt to shrink the footprint of the system; (e) a similar design as the current device but with the refocusing actuator on the side.

perturbations. Third, *clinical operability:* the module should integrate at the camera port, preserve parfocality, minimise alignment burden, and be practical for non-specialist clinicians to operate in routine settings. These goals are referenced throughout Chapters 3 and 4 to justify design decisions.

Guided by these goals, Figure 3.1 presents the resulting device as both a photograph and a CAD model, with component labels corresponding to the section headings in this chapter. The performance of this prototype is characterised in Chapter 5, and it is used for the autofocus applications described in Chapter 6 unless otherwise noted. Figure 3.2 shows earlier prototypes, illustrating the evolution of the optomechanical layout and component choices that led to the final design.

3.2 Optomechanical principle

The fundamental refocusing principle, first conceived by Paul Meyer, is demonstrated in the optical diagram shown in Figure 3.3. The depth-sensing component is shown inset, where case (i) represents a focused imaging camera, case (ii) represents loss of focus due to increasing object distance, and (iii) represents loss of focus due to decreasing object distance.

In words, the autofocus device is interposed in the parallel beam of any infinity-corrected optical system; the most common system for optical microscopy in current use. The parallel beam carrying the image is divided into primary and secondary beams by a 90:10 beam splitter. Ninety percent of the photons are converged onto the primary imaging camera, which is positioned conventionally and provides the microscope image to be studied. Ten percent of the photons are converged onto a secondary camera sensor that is tilted by an angle θ with respect to the microscope's image plane; this camera is referred to as the "tilted sensor" hereafter. The tilt angle means that one edge of the tilted sensor samples a nearer focal plane, while the opposite side samples a farther one. Therefore, the microscope's image plane is represented by a location of best focus (LoBF) on the tilted sensor, with increasing blur on either side of that point, as shown in Figure 3.3c. Changes in the object distance due to movement artefacts move the LoBF across the sensor at right angles to its axis of tilt, as shown in the sample images in Figure 3.3c. When imaging begins, the desired focal plane for the primary camera is selected by the operator (often the centre of the tilted sensor), and the corresponding LoBF on the tilted sensor is recorded. At any moment, the direction and distance the band has moved from its starting position, $e(t)$, informs the change in optical power required to restore a sharp primary image.

Refocusing is undertaken by an adjustable-focus lens, placed upstream of the beam splitter. The air-lens consists of nested plano-convex and plano-concave elements of equal focal length and glass type, and was originally conceived by Paul Meyer. The lighter plano-convex element is mounted and driven along the optical axis by a piezoelectric actuator, which varies the air-lens separation d and thus the optical power $P_{\text{air}}(d)$. The air-lens has no optical power when the lenses are apposed, but develops increasing positive power as the airgap between them is enlarged. Crucially, the air-lens causes almost no change in image magnification as the airgap increases; this is demonstrated by performance characterization in Chapter 5.

To formalise the behaviour of the tilted sensor and air-lens, we first fix a consistent coordinate system and sign convention. In object-space, distances along the optical axis increase from the objective toward the specimen; positive axial motion of the specimen increases object distance, whereas negative motion decreases object distance. In image-space, positive axial motion of the specimen correspondingly causes the image beam to be focused

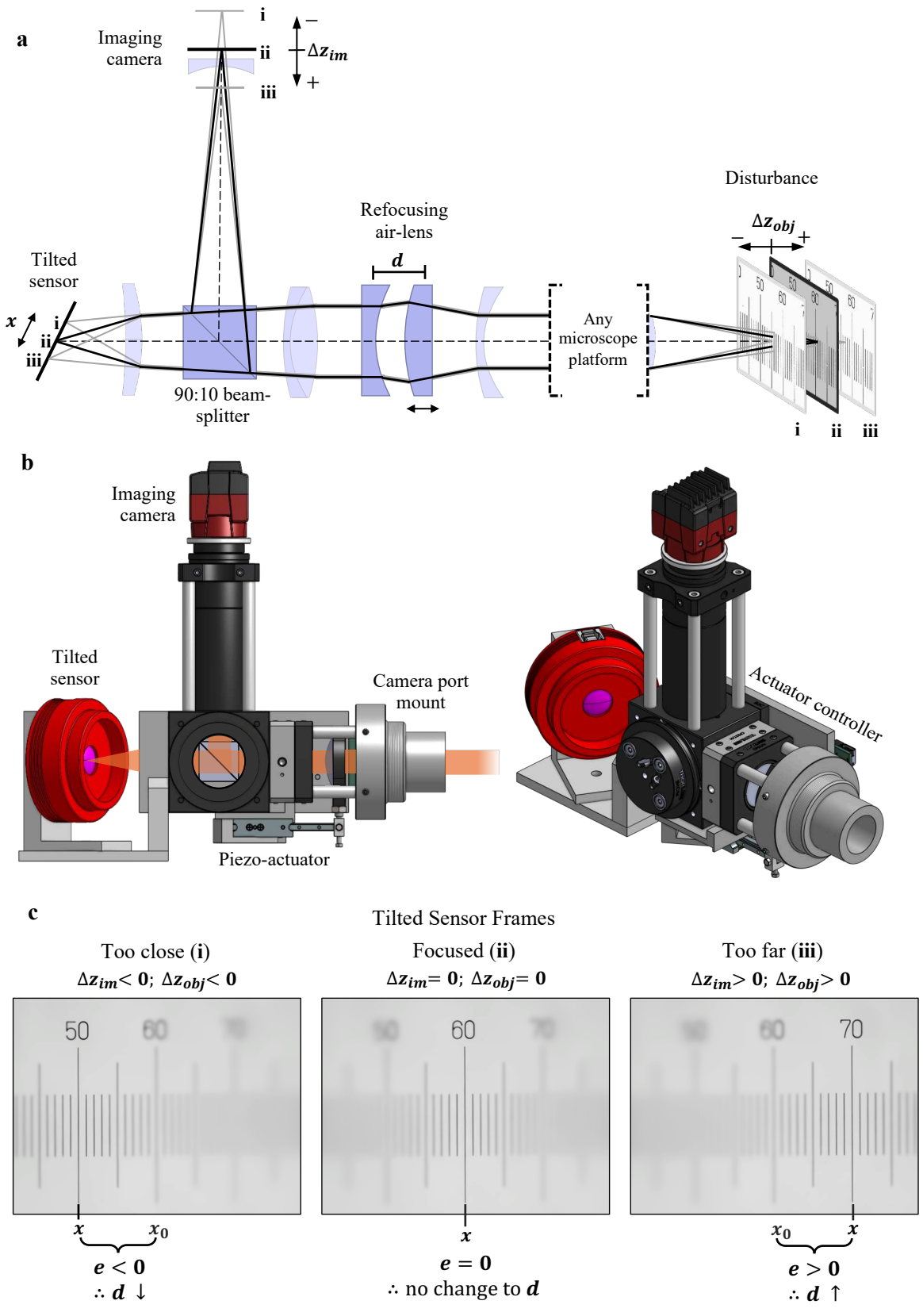


Fig. 3.3 (a) An optical ray diagram demonstrating the core refocusing principle. (b) A CAD model showing mechanical implementation and beam-path. (c) A tilted sensor image when the target is too close, focused, and too far.

in front of the imaging camera, whereas negative motion causes the nominal image plane to shift behind the imaging camera. To relate object-space motion to image-space shift of the nominal image plane, we use the approximation [6],

$$\Delta z_{\text{im}} \approx M^2 \Delta z_{\text{obj}}, \tag{3.1}$$

where Δz_{obj} is the signed object-space displacement of the specimen along the optical axis, Δz_{im} is the signed image-space displacement of the nominal image plane, and M is the lateral magnification of the objective–tube lens pair [6]. For an infinity-corrected microscope,

$$M = \frac{f_T}{f_o},$$

where f_T is the focal length of the tube lens and f_o is the focal length of the objective [6]. Equation (3.1) is the well-known longitudinal magnification relation from first-order (Gaussian) optics. Near best focus it predicts that image-space defocus scales as M^2 times the object-space defocus.

Equation 3.1 and the following relations are based on several common assumptions appropriate for first-order approximations of Gaussian optics [6]. First, the optical system is treated with paraxial and thin-element approximations, such that rays make small angles with the optical axis and higher-order aberrations or field curvature are neglected. Second, the air-lens is assumed to modify image focus with negligible change in lateral magnification over its operating range, as verified experimentally in Chapter 5. Finally, defocus excursions are assumed to remain within the range over which the sharpness metric used to determine the LoBF varies monotonically and approximately linearly with defocus. These assumptions are the standard starting point in first-order imaging. The mathematical models produced by these assumptions are verified empirically in Figure 3.4 and in the following section, with only minor deviations between theory and observation.

With assumptions identified and with an established object-space and image-space relationship, we now aim to model the mathematical relationship between three variables: the change in object distance of the imaging target Δz_{obj}, the air-lens separation d, and the LoBF along the tilted sensor x. To form these models, we vary one variable and measure the outcome of another, while holding the third variable constant. Each mathematical model provides expected behaviour for the air-lens, tilted sensor, or both, and this expected behaviour is verified with empirical data shown in Figure 3.4.

Case 1: fixed air-lens separation (d constant). To develop a formal model of the tilted sensor, we assume Δz_{obj} may vary while d is held fixed. This allows us to model how a shift

of the image plane by Δz_{im} is transformed by a sensor tilted by an angle θ into a lateral displacement x of the LoBF across the sensor. Let the nominal image plane be $z = 0$ and the tilted sensor be the plane

$$z \;=\; x \sin\theta, \tag{3.2}$$

where x measures distance along the sensor (for our device, along the horizontal). The sensor's horizontal projection in the microscope's image plane $u = x\cos\theta$, so equivalently $z = u\tan\theta$. The LoBF on the tilted sensor is given by the intersection of the defocused image plane $z = \Delta z_{\text{im}}$ with Eq. (3.2), which yields

$$x \;=\; \frac{\Delta z_{\text{im}}}{\sin\theta}. \tag{3.3}$$

Converting to pixel units with sensor pixel pitch p (m px^{-1}) and Δz_{im} in metres, and introducing x_0 as the pixel coordinate of the LoBF when the imaging camera is in focus, we obtain the operational mapping

$$x \;=\; x_0 \;+\; \frac{1}{p\sin\theta}\,\Delta z_{\text{im}}. \tag{3.4}$$

Combining Eqs. (3.1) and (3.4) gives the direct relation between object-space motion Δz_{obj} and LoBF position x,

$$x \;=\; x_0 \;+\; \frac{M^2}{p\sin\theta}\,\Delta z_{\text{obj}}. \tag{3.5}$$

Equivalently, in terms of the error signal $e = x - x_0$ (in pixels),

$$e = \frac{M^2}{p\sin\theta}\,\Delta z_{\text{obj}}. \tag{3.6}$$

Equations (3.4) and (3.5) explain the first empirical calibration data shown in Figure 3.4A: when the air-lens separation is held constant and the object distance is swept, the LoBF moves approximately linearly across the tilted sensor, with a slope that scales as M^2. In practice, the measured curve is close to linear over the device's intended operating range near the centre, with mild curvature and eventual saturation as the LoBF approaches the sensor edges. This behaviour reflects both the finite numerical aperture (which softens sharpness extrema) and the fact that the actual nominal image plane becomes a truncated ridge when the intersection falls outside the width of the tilted sensor.

Next, to develop a formal model of the air-lens, we use the first-order thin-lens approximation for each refracting element of the nested pair with paraxial power $P_i = 1/f_i$, separated by a distance d. Using Gullstrand's equation for separated thin lenses, the equivalent power

of the pair is

$$P_{\text{eq}}(d) \;=\; P_1 + P_2 - d\,P_1 P_2, \tag{3.7}$$

which is accurate to first order in the separation [6]. In our implementation, the elements have equal focal lengths f_0 of opposite sign, $f_1 = f_0$ and $f_2 = -f_0$, so $P_1 = +1/f_0$ and $P_2 = -1/f_0$. Substituting into Eq. (3.7) gives

$$P_{\text{air}}(d) \;=\; \frac{d}{f_0^2}, \tag{3.8}$$

showing that, within the paraxial approximation, the optical power of the air-lens increases linearly with its mechanical separation d.

Because the air-lens is located in the collimated section between the objective and the tube lens, this additional optical power slightly perturbs the effective focal length of the imaging relay. For collimated input, the nominal image plane lies one tube-lens focal length f_T behind the tube lens. With the added power from the air-lens $\Delta P = P_{\text{air}}(d)$, the effective focal length becomes $f_T' = 1/(P_T + \Delta P)$, where the power of the tube lens $P_T = 1/f_T$. For $|\Delta P| \ll P_T$, a first-order approximate expansion gives

$$\Delta z_{\text{im}}(d) \;\equiv\; f_T' - f_T \;\approx\; -f_T^2\,\Delta P \;=\; -\frac{f_T^2}{f_0^2}\,d. \tag{3.9}$$

Equation (3.9) shows that the air-lens produces a nearly linear shift of the imaging plane with separation d.

Case 2: LoBF held constant ($x = x_0$). Closed-loop autofocus is achieved by choosing x_0 as a setpoint and adjusting d so that the measured LoBF x matches x_0 despite specimen motion. In small-signal form we require the sum of object-induced and lens-induced image-plane shifts to vanish,

$$\Delta z_{\text{im}}(\text{object}) \;+\; \Delta z_{\text{im}}(d) \;=\; 0.$$

Using Eqs. (3.1) and (3.9) solving gives

$$d \;=\; \frac{f_0^2}{f_T^2}\,M^2\,\Delta z_{\text{obj}}. \tag{3.10}$$

Equation (3.10) explains the empirical calibration data shown in Figure 3.4B: when the LoBF is held constant and the object is moved axially, the required air-lens separation varies linearly with object distance, and the slope scales with M^2 across objectives. Intuitively,

higher magnification increases longitudinal magnification, so more optical power change is needed to null the same object-space perturbation.

Case 3: object distance held constant ($\Delta z_{\mathrm{obj}} = 0$). Combining Eq. (3.9) with Eq. (3.4) yields the mapping from air-lens separation to LoBF displacement on the tilted sensor for a stationary specimen,

$$e(d) \;=\; -\,\frac{f_T^2}{p\,\sin\theta\,f_0^2}\,d. \tag{3.11}$$

This relation explains the empirical calibration data shown in Figure 3.4C: when the specimen remains fixed and the air-lens separation is varied, the LoBF shifts approximately linearly across the tilted sensor with a slope that depends only on the optical configuration of the air-lens and tube lens. Crucially, Eq. (3.11) contains no dependence on objective magnification M. This agrees with measurements showing that the pixel-to-actuator sensitivity is essentially identical for all objectives tested, with slight deviations due to the first-order assumptions made. This invariance is operationally important, since it means that the controller gain factor $K_{ed} = \partial e/\partial d$ is determined solely by the air-lens and relay optics and does not require recalibration when objectives are exchanged.

Together, Eqs. (3.5), (3.10), and (3.11) provide a complete first-order description of the system: they describe the three-way correspondence between object position, air-lens actuation, and LoBF displacement on the tilted sensor that underpins closed-loop autofocus operation.

The three relations predict the results of the three calibration experiments. First, with fixed d and object distance swept in 0.1mm increments, the LoBF moves linearly in the central region, becoming non-linear toward the edges. Second, with the LoBF held constant and the object distance swept, the demanded d varies linearly (Fig. 3.4B). Third, with the object fixed and d swept, the LoBF moves approximately linearly with a slope that is nearly invariant across objectives (Fig. 3.4C). Minor differences in slope linearity here demonstrate the error due to the first-order assumptions listed earlier, but are managed by gain-scheduling as discussed in Section 4.4.1. In all three cases the empirical curves show agreement with these first-order predictions over the working range.

Finally, it is helpful to make the control-relevant sensitivities explicit. Differentiating Eqs. (3.5) and (3.11), we obtain

$$K_{xz} \;\equiv\; \frac{\partial x}{\partial \Delta z_{\mathrm{obj}}} \;=\; \frac{M^2}{p\,\sin\theta}, \qquad K_{xd} \;\equiv\; \frac{\partial x}{\partial d} \;=\; -\,\frac{f_T^2}{p\,\sin\theta\,f_0^2}.$$

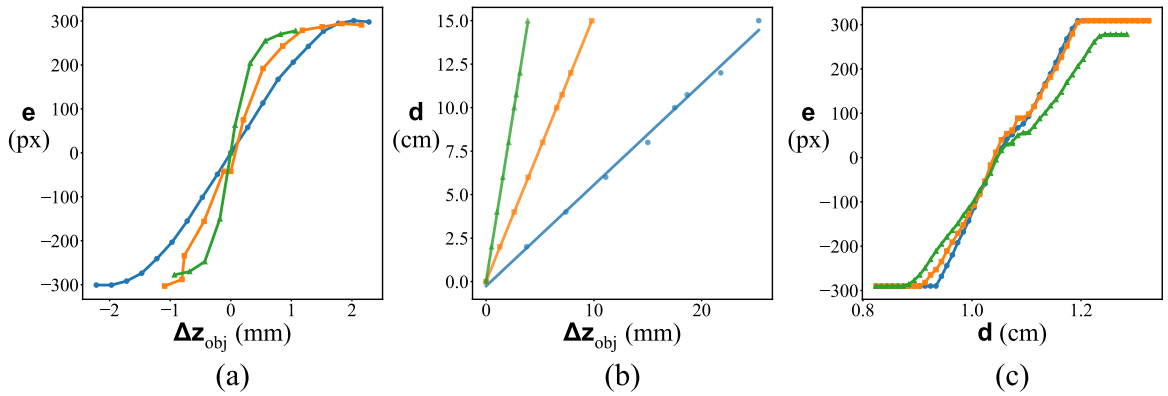


Fig. 3.4 Empirical calibration data relating Δz_{obj}, e, and d; blue is 12x, orange is 20x, and green is 32x magnification. (a) Air-lens separation held constant while object distance swept, LoBF recorded; (b) LoBF held constant while object distance swept, air-lens separation recorded; (c) object distance held constant, air-lens separation swept and LoBF recorded.

Thus the measurement gain K_{xz} increases quadratically with magnification, while the actuation gain K_{xd} is independent of magnification. A proportional–derivative controller acting on the pixel-domain error $e(t) = x(t) - x_0$ and commanding d therefore sees essentially the same plant sensitivity from actuator to sensor across objectives; only the mapping from object disturbance to measured error changes with M. Empirically, this is why the same controller gains work robustly across different objectives. Because the actuator–sensor path is magnification-invariant, loop stability and bandwidth are dominated by the air-lens mechanics and the sensor/compute latency, not by the objective choice.

Finally, the decision to reside entirely in the collimated section of an infinity-corrected microscope underpins the universality and stability of the module. In infinity systems the objective projects the specimen as a parallel beam and the tube lens performs the final imaging; auxiliary optics can be inserted into this “infinity space” by way of a camera port. This design pattern is widely adopted in modern research microscopes and is emphasised in standard texts on microscope optics and confocal systems [41, 88]. However, in Section 6.3, we do demonstrate adaptation to a convergent camera port benchtop microscope, by adding an additional optical element upstream of the autofocus device.

In summary, the optomechanical architecture can be approximated by a set of near-linear relations linking object motion, lens separation, and LoBF position. The empirical slopes align with the mathematical models, with some variance due to first-order assumptions. Because the actuator–sensor path is magnification-invariant, this optomechanical principle generalises to any objective magnification on the same microscope without any modification. Then, by adjusting a controller gain constant, which is made possible by the UI demonstrated in Section 4.6, the device adapts to any other infinity-corrected microscope. Even further, with

an optical element added upstream, the device operates successfully with convergent camera port microscopes as shown in Figure 6.10. Therefore, the optomechanical autofocus principle outlined in this section enables universal image-based autofocus for microscopy. Subsequent sections select specific components to realise this optomechanical design principle with sufficient accuracy and speed for video-rate operation in clinical and laboratory settings.

3.3 Imaging camera

Fig. 3.5 Potential imaging cameras procured and evaluated in bench-top tests.

The imaging camera selected for the final prototype in Figure 3.1 is the Allied Vision Alvium 1800 U-291m. The core requirements mirrored those of high-performance digital microscopy: sustained high frame rates ($\geq$120 fps), sufficient spatial resolution, and high quantum efficiency at the operating wavelengths. Application-specific constraints further included (i) a global shutter to avoid rolling-skew artefacts during rapid saccades, (ii) a monochrome sensor to maximise sensitivity and simplify the optical train, and (iii) strong responsivity in the 505-575 nm band, which is strongly absorbed by hemoglobin. The U-291m integrates Sony's IMX421 global-shutter CMOS sensor (2/3" format, 2.9 MP at 1944 $\times$ 1472, 4.5 µm pixels) with a full-frame rate of 144 fps and a peak quantum efficiency of $\approx$ 73 % at 529 nm, meeting these criteria. In addition, the USB 3.0 interface and actively supported software development kit (SDK) simplified integration with our acquisition software and helped ensure long-term compatibility.

Candidate devices were first screened via technical datasheet review. For example, several Basler ace2 models (notably the a2A1920-160umBAS) were considered but deprioritised due to availability constraints and resulting long lead times. Three principal candidates were ultimately procured for laboratory testing, which are shown in Figure 3.5 and compared side-by-side in Table 3.1. Early prototypes employed a Prosilica GC1380H (Sony ICX285, 2/3", 6.45 µm), which delivered the characteristic low noise and image quality of CCD sensors but was limited to 30 fps. As a discontinued model, it also raised concerns about

ongoing software support. The Alvium 1800 U-240m (Sony IMX392, 1/2.3", 3.45 µm) superseded the Prosilica in subsequent iterations, enabling much higher frame rates (up to 173 fps) with a modern SDK. However, its smaller optical format reduced field-of-view coverage and per-pixel photon capacity, yielding lower image quality in side-by-side tests. By contrast, the Alvium 1800 U-291m matched the per-frame image quality of the Prosilica while retaining the high-rate acquisition and software support of the Alvium line.

Table 3.1 Comparison of potential imaging cameras.

Property	**Prosilica GC1380H**	**Alvium 1800 U-240m**	**Alvium 1800 U-291m**
Sensor chip	Sony ICX285	Sony IMX392	Sony IMX421
Sensor type	CCD	CMOS	CMOS
Optical format	2/3"	1/2.3"	2/3"
Resolution	1360 × 1024 (1.4 MP)	1936 × 1216 (2.4 MP)	1944 × 1472 (2.9 MP)
Pixel size (µm)	6.45	3.45	4.5
Max frame rate (full)	30 fps	∼173 fps	∼144 fps
Shutter type	Global	Global	Global
QE @ 525–530 nm	54 %	64 %	73 %
Interface	GigE	USB 3.0	USB 3.0
Notes	Large sensor; low fps; discontinued	Small sensor; high fps; low cost	Large sensor; high fps; higher cost

In summary, the Alvium U-291m provides a favourable balance of speed, image quality, and software operability. Its global shutter prevents skew during saccades, and the relatively large 2/3" format with 4.5 µm pixels increases the per-frame photon budget and reduces noise. Finally, its monochrome design and strong green sensitivity align with the hemoglobin absorption band exploited during HVI. This camera was therefore instrumental in achieving the design goals of speed, accuracy/robustness, and clinical operability.

3.4 Tilted sensor

The tilted sensor used in the final prototype shown in Figure 3.1 is the ZWO ASI 120MM-S (USB 3.0). It employs an ON Semiconductor AR0130CS CMOS sensor (1/3" format, 1280 × 960 pixels, 3.75 µm pitch). The requirements for selecting the tilted sensor were more stringent than for selecting the imaging camera. The sensor had to operate in low light

Fig. 3.6 Potential tilted sensor options procured and evaluated in bench-top tests.

conditions, as only about ten percent of the beam is provided following the 90:10 beam splitter. It also had to sustain video-rate acquisition (60 Hz+) for closed-loop control, and produce a stable image despite being mounted at an oblique angle to the image plane. A monochrome sensor was another requirement, to maximise sensitivity and simplify optics. Sharing a software SDK with the imaging camera was also desired to reduce integration overheads, but this was not a necessity.

Many candidates were considered with these requirements in mind. Figure 3.6 shows the camera models that were explicitly compared in bench-top laboratory experiments. A number of additional cameras not shown in Figure 3.6 were reviewed at the datasheet level for completeness (e.g., the Altair 174M) , but were found to be inferior in crucial areas.

Early prototypes employed the older USB 2.0 model of the ZWO ASI 120MM, an astrophotography-oriented camera with strong low-light behaviour but a 30 fps ceiling imposed partly by the bus. The USB 3.0 successor, the ASI 120MM-S, promised similar imaging characteristics while enabling higher frame rates. The USB 3.0 interface (≈4,800 Mb/s versus 480 Mb/s for USB 2.0) sustained 60 fps acquisition while preserving the low-light performance observed in the USB 2.0 model, and it was inexpensive and compact. We also evaluated Allied Vision's Alvium 1800 U-240m, because matching manufacturers for the imaging camera and tilted sensor would allow unification behind a single SDK, and the larger sensor could increase the usable detector range. However, in practice the Alvium's C-mount had a pronounced front lip which cast a shadow on the sensor at the required obliquity, and its length restricted the permissible geometry. Therefore, we test the board-level variant of the Alvium 1800 U-240m directly against the ASI 120MM-S in Figure 3.7.

Benchtop experiments mounted the ASI and Alvium candidates on a rotary stage and swept the tilt from 0° to 90° in 5° increments relative to the plane orthogonal to the optical path, recording per-tilt intensity and image sharpness using OpenCV routines. A sampling of images from this benchtop experiment are displayed in Figure 3.7. At 20° the Alvium's mean

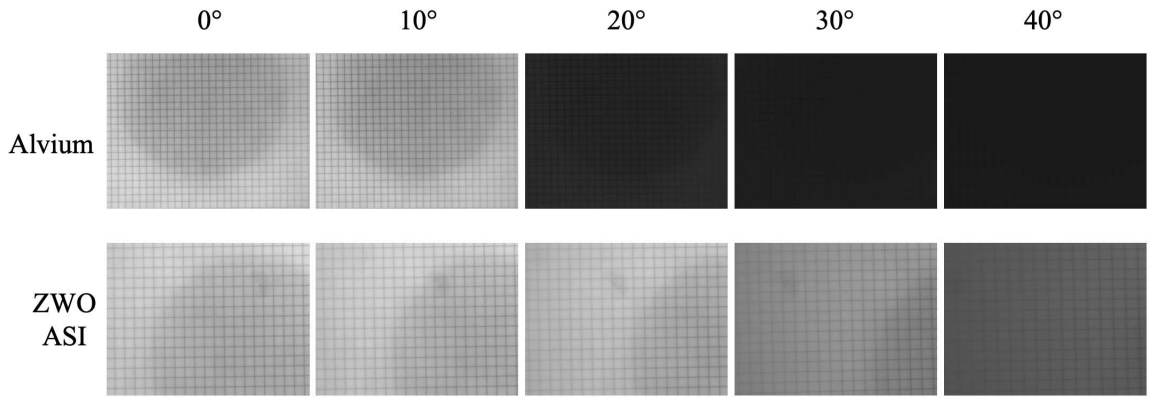


Fig. 3.7 Sample tilted sensor images at various tilt angles.

pixel intensity fell sharply to 35 on an 8-bit scale (n = 3, max = 255), whereas the ZWO remained around 150. Beyond 20° only the ZWO camera delivered sufficient signal-to-noise to extract reliable sharpness metrics for the grid imaging target. Because our prototypes operate with a 25° sensor tilt, these results effectively screened out the Alvium options for the tilted sensor role. We therefore selected the ZWO ASI 120MM-S as the tilted sensor. Overall, the ASI 120MM-S USB 3.0 camera provided the necessary combination of speed, sensitivity, and compactness to achieve reliable, accurate LoBF tracking when inclined relative to the microscope's image plane.

Finally, the tilt angle in the final prototype was determined by considering the trade-off implied by Eq. (3.5). Since the LoBF slope scales as $1/\sin\theta$, increasing θ improves pixel-domain sensitivity to axial motion but reduces the linear working range before the LoBF ridge reaches the sensor edges. For all the prototypes in this thesis, we chose $\theta = 25°$, chosen to prioritise the successful development of a robust, video-rate prototype that can manage all blinks and microsaccades. As the control loop matured, most LoBF excursions concentrated near the sensor centre, suggesting that smaller tilts may be optimal in later prototype revisions. This is discussed further in Chapter 7.

3.5 Refocusing actuator

In the final prototype shown in Figure 3.1, the plano-convex element of the tunable air-lens is translated by an Xeryon XLA-5-65-1250 piezomotor. On the back of the device, as shown in Figure 3.1C, is the Xeryon XD-OEM controller which provides power and a closed-loop serial interface. This mechanism was chosen as the refocusing action because it met our precision and speed requirements while integrating cleanly with the enclosure and optics.

Thorlabs ELL17/M Xeryon XLA-5-65-1250

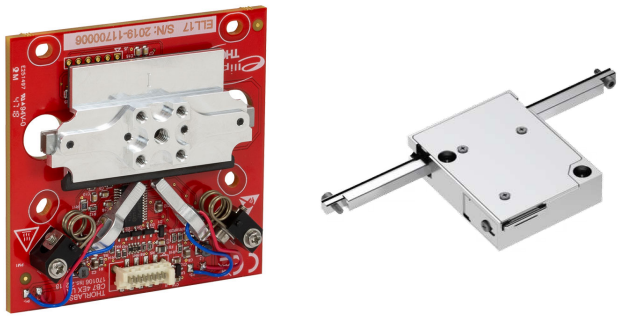


Fig. 3.8 Potential actuators procured and evaluated in bench-top tests.

Before settling on a tunable air-lens, Paul Meyer evaluated a commercial liquid lens (Optotune) for refocusing, but noticed that the gravitationally induced astigmatism of its refractive surface introduced unacceptable deformation of the microscope image. We therefore retained the air-lens concept, which enables high-speed refocusing in the infinity-corrected beam with negligible change in magnification. At the datasheet level we compared classes of actuators, including a voice-coil linear stage, a miniature stepper on a lead screw, a compact DC servo with encoder, and a hydraulic micro-actuator. None offered the combined speed, repeatability, stiffness, and compactness found in ultrasonic linear piezomotors. The two piezo-motors shown in Figure 3.8 promised high speed and fine control of a ~100 g lens in a compact package.

The refocusing actuator had to satisfy several requirements. It needed to support video-rate operation with minimal added latency, deliver sufficient stroke range to span the autofocus capture range, and provide high positional precision and repeatability. Equally important was continuous operation; the video-rate control loop may update lens position up to 60 times per second during typical ocular motion, so any explicit duty-cycle limit would throttle throughput. Mechanical form factor, mass, and cost also mattered because the lens must be mounted in the infinity-corrected space without inducing tilt or vibration into the microscope.

Our first prototype used Thorlabs' ELL17K/M piezo linear stage (Figure 3.8). To quantify its effect on the control loop, we commanded a train of ±1 mm moves while filming the lens. Frame-by-frame analysis showed a peak translational speed of about 180 mm/s, in agreement with the manual, which is more than adequate for live autofocus. However, we observed an unexpected pause between commands that scaled with the requested travel, evidently enforcing an internal duty-cycle limit near 40 percent. This inter-move delay, rather than the move itself, dominated the latency between a detected movement artifact and subsequent

Table 3.2 Comparison of the two refocusing actuators used in this work.

Property	**Thorlabs ELL17K/M**	**Xeryon XLA-5-65-1250**
Travel (stroke)	28 mm	30 mm
Max velocity (no load)	180 mm/s	400 mm/s
Positioning resolution	0.98 µm	1.25 µm
Backlash	n/a	±1 µm
Acceleration (no load)	6.0 m/s^2	730 m/s^2
Duty cycle	40%	continuous
Nominal Payload	200 g	170 g
Lifetime	>100 km	>1000 km
Dimensions (body)	68×67×15.3 mm	38×30×9.1 mm
Encoder type	magnetic	optical

accommodation. In practice, we found the Thorlabs piezomotor could only update position up to 4.3 Hz (n=3) for even ±0.1 mm moves, which was unacceptable. After extended correspondence, Thorlabs support disclosed a non-documented "dynamic delay" parameter, which by changing from 0*vv*05 to 0*vv*30 brought this latency up to 8.6 Hz (n = 2). However, Thorlabs also cautioned that higher dynamic delay settings increase the risk of overheating the resonant piezo paddles. Therefore, we evaluated a Xeryon XLA-5-65-1250 actuator as a replacement, with the key parameters shown in Table 3.2.

Specification by specification, the XLA-5 exceeded the ELL17K/M in almost every metric relevant to this application. It offered more than double the maximum velocity and over two orders of magnitude higher acceleration, enabling rapid refocus without saturating the control bandwidth. Its continuous-duty rating eliminated the enforced pauses that limited the Thorlabs unit, and its optical encoder provided more stable position feedback at high speeds. The rated lifetime of over 1000 km ensured durability under continuous 60 Hz operation. Finally, the compact form factor simplified mechanical integration, although the necessity of the XD-OEM controller was a challenge to mount, solved by Paul Meyer. Recently, Xeryon has released a model with an integrated controller, which is mentioned in Chapter 7 as a clear way to improve on the existing prototype.

We therefore transitioned to the XLA-5 actuator. Unfortunately, this was not a straightforward transition: out of the box the Xeryon actuator introduced visible lens jitter linked to resonant modes of its rod-motor assembly. This jitter meant blurring of the microscope image which was unacceptable, especially since such jitter wasn't an issue with the more stable ELL17K/M. Fortunately, the XD-OEM controller provided much finer control over drive frequency, internal gain constants, and related parameters. Through bench tuning

and consultations with Xeryon engineers, we identified a band of drive frequencies and corresponding gains that suppressed the resonant jitter to negligible levels. Another issue was orientation-dependence in the actuator behaviour, which remains even after careful tuning of control parameters and consulting support engineers. But the behaviour differences of the XLA-5 when vertical or horizontal are small, and can be absorbed by mechanical bracing and minor adjustments to the outer-loop gains using the user interface, as discussed in Section 4.6.

Following this development work, the XLA-5 delivered stable, fast, and precise continuous refocusing. Repeating the ±0.1 mm benchtop test, it executed commands at 27.3 Hz (n=3), compared to 8.6 Hz (n=3) for the ELL17K/M. At its minimum encoder step of ±1.25 µm, the actuator could execute move commands at 60 Hz and above without loss of accuracy. Its dynamic performance is characterised further in Chapter 5, together with the measurement and control loop. In sum, the XLA-5 achieved the required throughput without thermal or duty-cycle constraints, meeting the speed, precision, and robustness targets for a video-rate autofocus module.

3.6 Computing

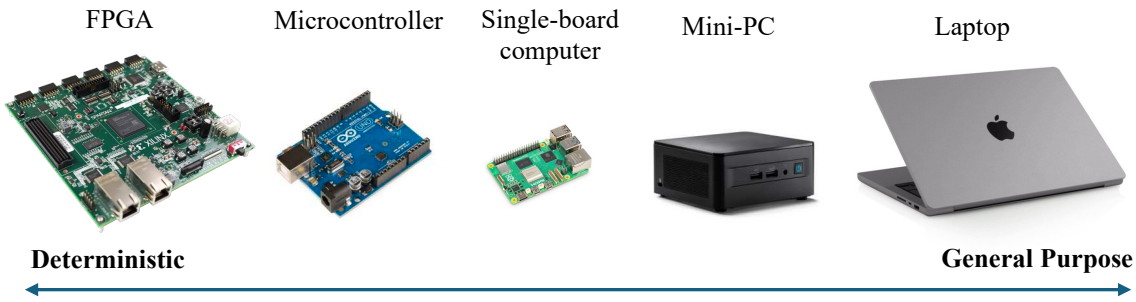


Fig. 3.9 Considered computing options.

The final prototype shown in Figure 3.1 is operated using an Intel NUC mini-PC, with 16GB of RAM and a 12th generation i3 CPU from Intel. The second autofocus prototype operating in the Clifford Albutt lab uses a similar Intel NUC mini-PC, but with 32GB of RAM and a 13th generation i3 CPU from Intel. These CPU and RAM specifications are typical of mid-range laptops and computers today, meaning the autofocus device operates using readily-available commercial hardware. Both devices also come with three USB 3.0 ports, two USB-C ports, and two HDMI ports, which are crucial for minimising latency and simplifying device control.

When selecting a computing platform for real-time autofocus with a tilted sensor, the requirements are straightforward but non-trivial. The system must ingest two high-bandwidth camera streams, compute sharpness and registration metrics simultaneously at video rate, and close the control loop with a piezo-driven air-lens, all while presenting an operator interface suitable for clinical use. For reproducibility across sites, the development environment and drivers had to be stable, well supported, and easy to replicate.

It was useful to frame the available options along the axis shown in Figure 3.9, which runs from highly deterministic but specialised hardware to general-purpose systems with lower timing determinism. The most flexible options are laptops and desktops: they readily deliver the raw compute needed for image processing, but their form factor, built-in display, and peripheral I/O options make them ill-suited to permanent integration with a clinical instrument. A more compact and application-specific option are compact x86 or ARM "mini-PCs," including the Intel NUC. These preserve the advantages of a full operating system and drivers, but in a smaller footprint that can live with the clinical device, and with a wider range of I/O and display options. Single-board computers (SoCs) such as the Raspberry Pi occupy a still smaller and lower-power niche. However, in practice, SoC I/O bandwidth, CPU throughput, and camera driver support make support for two synchronised, high-speed streams unreliable. At the deterministic end are microcontrollers and FPGAs. They excel at low-latency control and fixed-function pipelines, but are mismatched to full-resolution video acquisition and to an algorithm that was still evolving. Microcontrollers simply cannot move or process the required data produced by high-bandwidth video streams. An FPGA implementation would have required custom hardware design and tooling before the processing pipeline had stabilised, inhibiting development speed.

A tempting alternative, not shown in Figure 3.9, is a GPU-accelerated platform such as the NVIDIA Jetson boards. GPUs offer massive data-parallel throughput and are attractive for convolutional metrics, feature extraction, and learning-based autofocus. In principle, a Jetson-class SoC could replace a mini-PC and add acceleration headroom for future deep-learning variants of the autofocus algorithm. In practice, however, the present autofocus workload is latency-sensitive and memory-bound rather than a throughput problem. In other words, the latency between movement detection and accommodation is minimised with the CPU's fast computation and minimal transfer overheads, rather than with a GPU's simple calculations performed in parallel. The Alvium and ZWO ASI USB3 cameras also lack true zero-copy paths into GPU memory, so each frame would pay an additional transfer latency cost with uncertain timing. Thermal and acoustic budgets also favoured a CPU-centric design; small, actively cooled GPUs can be audible and warm in a darkened exam room. For

these reasons, we deferred GPU adoption to a future revision aimed at higher frame rates, multi-camera fusion, or learning-based control. This discussion is revisited in Chapter 7.

Against this landscape, an Intel NUC mini-PC provided the best balance. It combined strong single-thread performance for latency-critical tasks with enough multicore capacity to run acquisition, computation, and actuation in parallel. The presence of multiple native USB ports simplified integration by allowing each component to connect directly. Previous prototypes had relied on intermediary USB hubs that introduced transfer latency. Unlike a laptop, the NUC could be mounted semi-permanently to the microscope without ergonomic penalties and without the variability introduced by battery management and sleep policies. Unlike an FPGA or microcontroller, it allowed us to iterate quickly in a conventional software environment as the focus metric and control logic evolved.

On the operating system side, we selected Linux for timing predictability and control over process-scheduling. Windows and MacOS often interleave background services that preempt user processes. While this is manageable for many applications, infrequent and unpredictable delays are problematic for closed-loop control at video rate. Linux also offered more transparent control of thread priorities and CPU affinity, and access to real-time scheduling classes. However, we prioritised platform-agnostic libraries like OpenCV for image processing and GTK for the interface, which kept the codebase OS-agnostic. Therefore, porting software to Windows or macOS remains feasible without significant architectural changes.

C++ was chosen as the software language to meet performance and control requirements that earlier prototyping in Python could not. While this was a significant development cost, C++ is reported to be around 10x faster than Python in benchmarks and in prior work [18]. C++ also enables multi-threading, explicit memory management, and careful avoidance of unnecessary processes in the critical control pipeline. Each of these features proved valuable in algorithm development. Further, C++ allowed tighter integration with vendor SDKs (e.g. those from Alvium and ASI) and precise control over error handling, which simplified maintaining performance over extended sessions. The overall software architecture is detailed in Chapter 4, and its measured performance is reported in Chapter 5.

In sum, an Intel NUC mini-PC running Linux with a C++ implementation best met the needs of a tilted-sensor autofocus system intended for clinical environments. It provided enough compute and I/O to run the dual-camera, closed-loop pipeline at video rate with predictable latency. It did so in a compact, quiet, and reproducible form factor. This choice of computing platform enables the device to achieve our design goals of speed, accuracy, and clinical operability.

3.7 Enclosure and interface

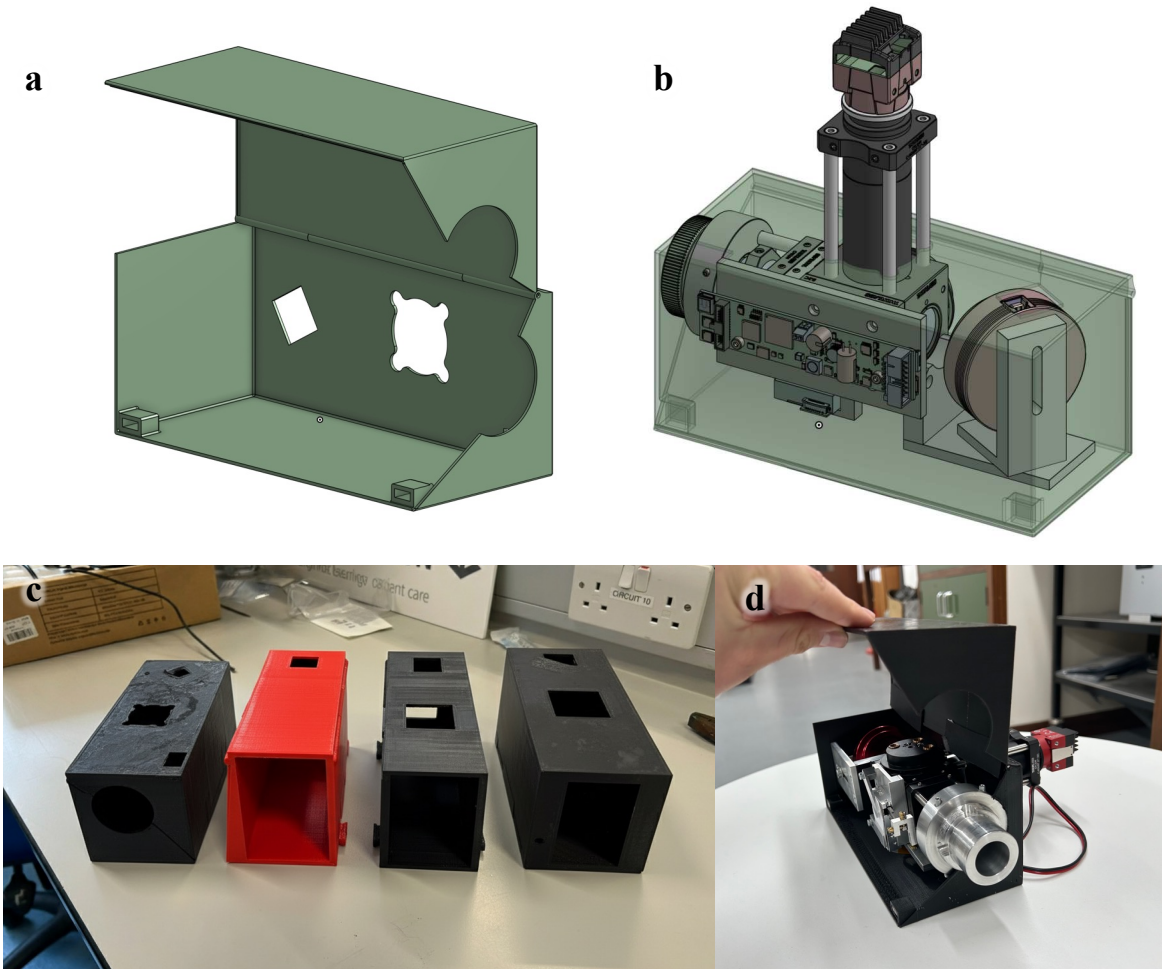


Fig. 3.10 (a) Enclosure in CAD with the hinge open. (b) Enclosure on the autofocus CAD model. (c) Four printed enclosures in total. (d) Final enclosure on the autofocus system.

Early prototypes were shielded from ambient illumination with improvised covers made from black cardboard or cloth. This typically constrained imaging to a darkened room, and left fragile assemblies exposed. To enable routine clinical operation in brightly-lit environments and to ease portability across microscopes and rooms, a light-tight, mechanically protective enclosure was required. Achieving an optically tight fit while preserving access for cabling and service demanded a scale-accurate CAD model of the autofocus device. This digital assembly captured device geometry for visualization (e.g. Figure 3.3b) and for designing cable pass-throughs, clearances, and magnet pockets in the enclosure before printing. Panels a–b of Figure 3.10 show the resulting enclosure model and its registration

to the device, while panels c-d illustrate the evolution from early prints to the final fitted assembly.

Four enclosures were fabricated in total, shown in Figure 3.10, each paired to a successive prototype of the autofocus device and incorporating lessons from the previous print. The first was a rectangular, top-down shell with a sliding bottom plate guided by opposing flanges, and with ports for the sensor and actuator leads. Although it established basic light-tightness, the sliding subassembly proved unreliable. Small flexures in either part allowed the plate to disengage, and the overall footprint was bulky. The second iteration replaced the sliding plate with a two-piece clamshell design, connected together with a hinge formed by a through-rod. The hinge allowed the enclosure to open for installation over the device and then close to form a light-tight seal, while remaining re-openable for later access. However, the hinge barrels printed undersised, and attempts to ream them for the rod introduced local warping and compromised alignment. A third print refined the hinge geometry and improved mechanical engagement, but an inadvertent change to orange filament produced excessive internal reflectivity. Lining the interior with black card restored adequate suppression of stray light and allowed use on the Clifford Allbutt laboratory prototype.

The final enclosure integrated the successful elements of previous iterations while introducing tighter tolerances and additional functional improvements. The hinge barrels were larger and more robust, while still providing an opaque surface on closing. The internal volume was minimised, and the cable passages for the tilted sensor and refocusing actuator were refined to achieve tighter tolerances. It also incorporated magnet compartments along the closing edge to provide a repeatable, “click” closure that simultaneously improved light sealing and sped assembly. In aggregate, the design delivered a robust, light-proof shell that protected the optics and electronics while permitting rapid access for adjustment and service.

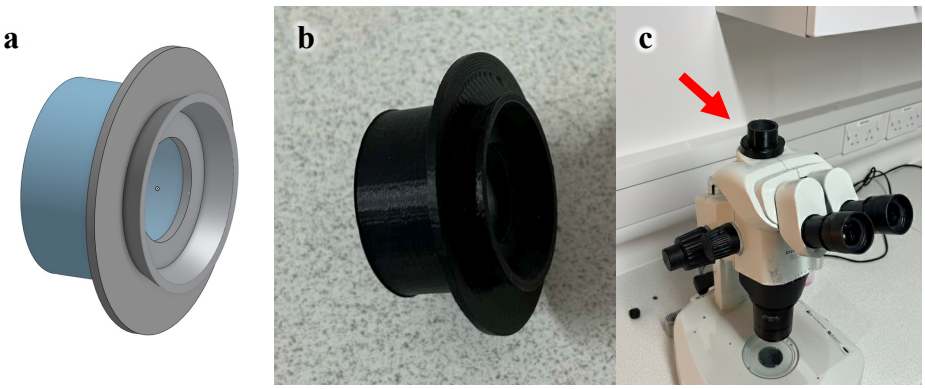


Fig. 3.11 (a) CAD model of the camera port interface. (b) Photo of the printed interface. (c) Interface on the Olympus SZX16 microscope.

Another additively-manufactured component was the interface required to mount the autofocus device on an Olympus SZX16 benchtop stereo-zoom microscope, as part of the generalization experiments described in Chapter 6. This adapter was modelled by referencing the geometry of an existing camera interface and by referencing the autofocus module's existing interface from the CAD assembly. The printed part achieved a functional fit on the first attempt; minor sanding established a firm interference fit at the module side. Critically, the adapter preserved access to the infinity-corrected path and provided space to insert an auxiliary optical element to compensate the SZX16's converging output beam. Because camera port standards vary across microscope manufacturers, this adapter serves as a template for rapid, low-cost interfaces to other platforms.

Together, the enclosure and camera-port interface are necessary to translate component-level choices into a deployable instrument. The enclosure improves robustness by shielding the optics from stray light and physical disturbance, and advances clinical operability by eliminating dark-room requirements and improving portability. The interface extends operability across optical platforms without compromising parfocality or alignment. Both elements thus support the chapter's design goal of preserving the speed and accuracy delivered by the sensing and actuation subsystems.

3.8 Conclusions

This chapter translated the requirements from Chapter 2 into a concrete, universal hardware platform for image-based autofocus, shown in Figure 3.1. We formalised the optomechanical principle that couples a tilted sensor to a tunable air-lens, and validated those first-order models with empirical calibration data. Then we discussed selection of the key components, including the imaging camera, tilted sensor, refocusing actuator, and computing platform, and evaluated component options through targeted bench tests. The final architecture comprises an Alvium 1800 U-291m imaging camera, a ZWO ASI 120MM-S tilted sensor, an XLA-5-65-1250 piezo actuator for refocusing, and an Intel NUC/Linux/C++ computing stack.

Considering the first design goal of *speed*, the piezomotor refocusing actuator and tilted sensor move the system from limited bursts of accommodation to genuine video-rate operation. The ELL17K/M was effectively capped at approximately 4.3–8.6 Hz for even the smallest moves, while the current XLA-5-65-1250 sustains approximately 27.3 Hz for $\pm$0.1 mm steps and $\geq$60 Hz at its minimum increment, collapsing refocus latency. In parallel, the ASI 120MM-S over USB 3.0 delivers stable 60 Hz LoBF tracking despite the 90:10 split and 25° tilt, doubling the 30 Hz ceiling of the earlier unit. The Intel NUC, running Linux

and C++, completes the speed-oriented design by providing a reproducible, low-latency computing platform with predictable thread scheduling and minimal I/O overhead.

Related to *accuracy and robustness* (Goal 2), the imaging chain and refocusing mechanics prioritise signal fidelity and repeatability. The Alvium U-291m's 2/3" IMX421 sensor with a global shutter, 4.5 µm pixels, and high 529 nm QE preserves detail and prevents saccadic skew. This improves on the earlier Prosilica GC1380H's 30 fps ceiling and the smaller-format U-240m. Upstream actuation via the air-lens changes optical power with negligible magnification shift, while the XLA-5's 1.25 µm resolution, >1000 km rated lifetime, and magnification-insensitive gain yields consistent plant behaviour across objectives and perturbations. Continuous actuation also improves robustness, allowing the XLA-5 to avoid instability by responding instantly to motion artifacts.

For *clinical operability* (Goal 3), the enclosure and computing platform translate a lab prototype into a deployable instrument. The light-tight, hinged, magnet-latched shell replaces ad-hoc cardboard/cloth shrouds, which enables bright-room use, protects cabling, and enhances portability. The printed adapter preserves alignment and demonstrates generalization to a variety of camera ports and microscopes. Finally, the Intel NUC provides a compact, reproducible host with ample I/O connectivity. Its predictable scheduling and consistent performance contrast with the variability of the earlier laptop-based system.

The system architecture outlined in this chapter provides a strong hardware foundation for universal, video-rate, image-based autofocus. Chapter 4 builds on this platform with a software acquisition and control pipeline that converts capable hardware into a reliable, video-rate autofocus system.

Chapter 4

Video-rate control and acquisition software

4.1 Introduction

This chapter presents the real-time control and acquisition software that enables accurate autofocus at or above video rate (≥ 60 Hz) using the hardware platform described previously. Central to this chapter is the sensing-to-actuation control loop which underpins the system's ability to achieve real-time autofocus; this represents a key contribution of this thesis. In addition, this chapter demonstrates real-time image registration, sensor fusion of the imaging camera and tilted sensor, and a user interface–each necessary for a clinical-grade refocusing platform. The complete codebase is available as open-source software on GitHub (https://github.com/lyncht248/Autofocus-Anything), which also hosts a detailed `README` with installation instructions and usage notes. The `README` also links to a hosted Doxygen-style documentation of each class and function.

Figure 4.1 presents a simplified software architecture and data-flow diagram of the codebase. The principal C++ classes appear in boxes and are colour-coded by function: components that interface with and manage the imaging camera appear in green, user-interface elements in blue, and the sensing-to-actuation autofocus subsystem in red. Solid arrows denote object ownership, while dashed arrows indicate data transfer and control dependencies. Hardware endpoints occupy the periphery of the diagram: the imaging camera is accessed through the vendor's `VimbaSDK`, the tilted sensor is operated by a custom `TiltedCam` class, and the air-lens actuator is driven via the vendor's `Xeryon` class. The figure is intentionally condensed for clarity and omits secondary pathways, several data relationships, and diagnostic utilities.

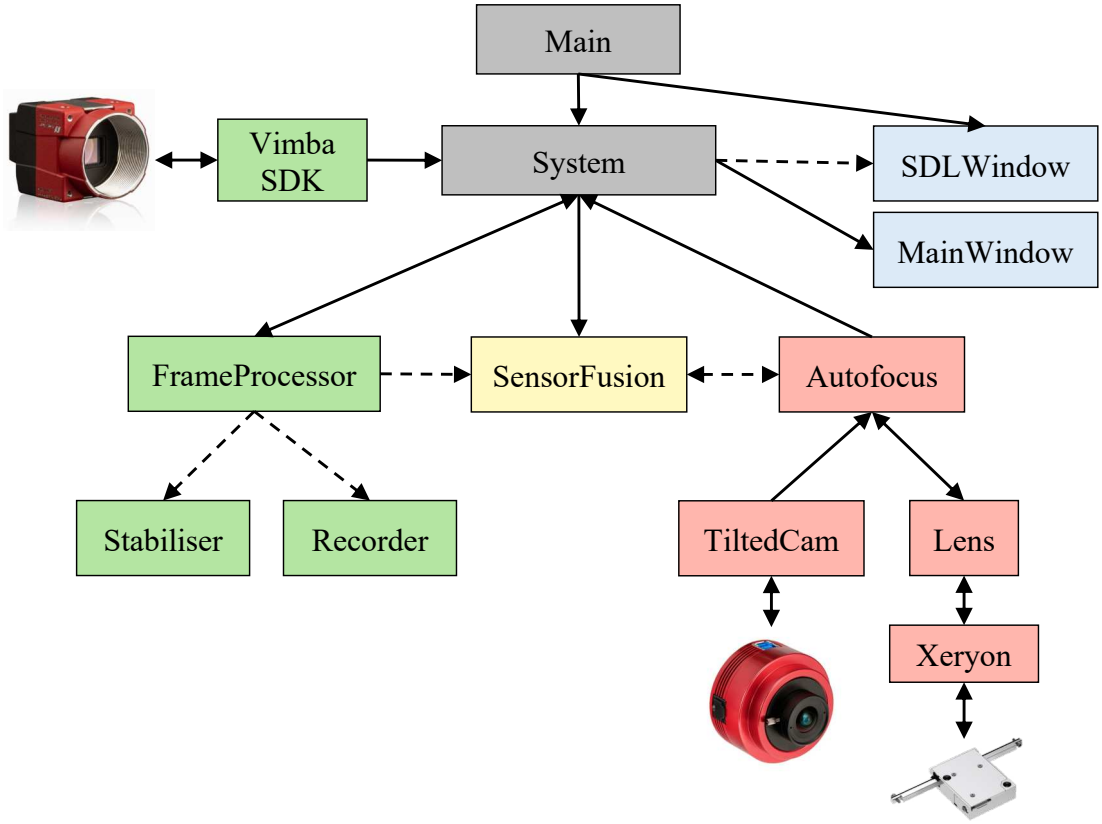


Fig. 4.1 Simplified software architecture and data-flow diagram. Green relates to the imaging camera, red to the tilted sensor, yellow to both sensors, and blue to the UI. Solid arrows denote object ownership, while dashed arrows indicate data transfer and control dependencies.

The codebase follows a standard central-controller pattern. The entry point `Main` launches the `System` class, which constructs, coordinates, and ultimately disposes of the principal services, including `FrameProcessor`, `SensorFusion`, `Autofocus`, `MainWindow`, and others not shown. Each class is defined by paired header (`.hpp`) and implementation (`.cpp`) files (e.g. `autofocus.hpp/.cpp`). The repository is organised in a conventional C++ layout with headers under `include/`, sources under `src/`, configuration files under `config/`, and unit tests maintained alongside the code; the exact file structure is documented in the `README` and not reproduced here. Helper classes common throughout the codebase include `Cond` for signalling, `Thread` and `TSQueue` for concurrency and queuing, `LogFile` and `NotificationCenter` for runtime logging and error reporting, and `Settings` for configuration.

With that architecture in mind, the chapter is organised as follows. Section 4.2 describes the concurrency model and the C++ techniques used to prioritise the autofocus pipeline

while maintaining simultaneous rendering, registration, and GUI responsiveness. It focuses on the acquisition, computation, and actuation processes of the core `Autofocus` class. Section 4.3 details the algorithm which computes the LoBF on the tilted sensor, including frame pre-processing, sharpness-curve computation, and the reduction to a robust LoBF value. Section 4.4 details the controller model used to drive the error between the desired and actual LoBF to zero. Section 4.5 summarises real-time image registration improvements required for real-time operation in the software application. Finally, Section 4.6 presents the desktop application that emerged through many design iterations with clinicians and researchers.

Throughout, the same design goals articulated in Chapter 3 motivate the engineering choices here. The first is *speed*, sufficient to sustain $\geq$60 Hz closed-loop control across the acquisition, computation, and actuation processes. The second is *accuracy and robustness*, to ensure reliable convergence across objectives and specimens and tolerance to artefacts such as brightness asymmetry and specularities. The third is *clinical-grade accessibility*, to integrate at the camera port, minimise alignment burden, and provide a user interface that makes the system practical for routine use by clinicians and researchers. These goals frame the development of a specimen-agnostic, objective-agnostic, video-rate autofocus system and the software that enables it.

4.2 Concurrency

The autofocus control loop comprises three processes: acquisition, computation, and actuation, which exchange image and state data at video rate. Figure 4.2 depicts these as parallel swim lanes, with sub-functions within each lane and arrows indicating data flow. The top panel (Figure 4.2A) illustrates the original sequential implementation, where a frame was fully acquired before any computation began, and the controller blocked while the actuator acknowledged motion. The bottom panel (Figure 4.2B) shows the concurrent design adopted in the current system: acquisition, LoBF computation, and actuator command proceed on separate threads with well-defined hand-off points, so that mechanical motion and rendering no longer stall image capture or focus estimation.

The sequential autofocus prototype from before the start of this thesis (Section 1.2) exemplified these limitations, as it was limited by both input/output and control blocking. Acquisition averaged 60.2 ms per frame ($\approx$16.6 Hz, $n = 4$), in part because each frame was written to and subsequently read from disk. Computation of the location of best focus (LoBF) required 27.5 ms on average ($\approx$36.2 Hz, $n = 2$). The occasional large-scale actuation would dominate the loop: a 1 mm command took 231.5 ms to complete, during which the software

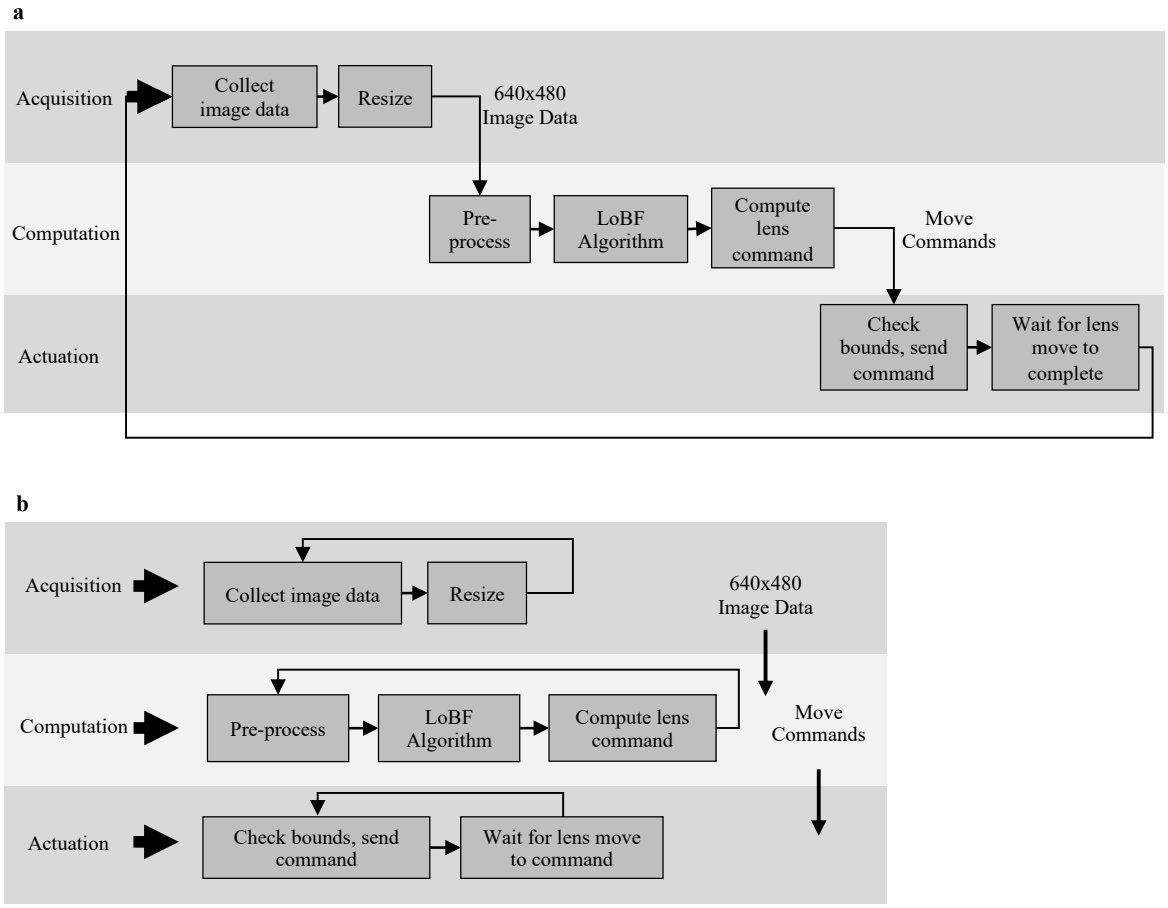


Fig. 4.2 Data-flow diagram of the acquisition, computation, and actuation processes with (a) a sequential architecture and (b) a concurrent architecture.

waited synchronously for a move-complete signal. During typical conjunctival imaging, these delays yielded a mean focus-adjust interval of 197.0 ms (≈5.1 Hz, $n = 3$), far below the target of video rate.

To overcome these limitations, the codebase presented in this chapter was newly implemented in C++, replacing all prior prototypes. C++ enables tighter control over memory, has native multithreading functionality, and can be up to 10 times faster for CPU-heavy tasks [18]. Several architectural changes accompanied the rewrite. Frames from the tilted sensor are retained in RAM and immediately discarded once the LoBF has been computed; disk writes are now optional for diagnostics rather than the default. On the actuation side, the software maintains an internal estimate of the lens position that is periodically reconciled with the hardware encoder, rather than querying position after every command. This halves serial traffic and removes a polling bottleneck from the critical path.

Concurrency is implemented with one thread per process, joined by thread-safe queues and condition variables. This design is inherently more complex than a sequential pipeline, as it requires explicit coordination to avoid race conditions. For example, the acquisition thread pushes frames to a bounded queue as they become available. The autofocus computation always requests the most recent fully received frame, a "latest-wins" policy that avoids race conditions while maximising temporal fidelity for control. In contrast, when every frame must be preserved, as in registration, the consumer drains the queue in order. All frames carry monotonic timestamps so that downstream modules can compute inter-arrival jitter, reject stale data, and correlate sensor and actuator events.

Scheduling policies are tuned to keep the autofocus loop on the CPU even when the desktop application is busy. The three critical threads shown run with elevated priority and CPU affinity so the operating system scheduler does not pre-empt acquisition during rendering bursts. On a typical workstation the application maintains between twelve and thirty active threads, for sensor I/O, autofocus, registration, display, recording, logging, and GUI responsiveness. Despite this concurrency, the design avoids oversubscription: compute-heavy tasks are bounded, and back-pressure behaviour is enforced by queue limits so that overload results in controlled frame dropping rather than unbounded latency.

With these changes, the control loop operates at video rate. For our prototype, acquisition occurs at 60 Hz (16.67 ms per frame). The complete LoBF computation, including pre-processing, sharpness-curve formation, and reduction, takes 1.34 ms end-to-end (details in Section 4.3), and is non-blocking. Actuation is also non-blocking; physical settling proceeds in parallel and is reconciled by the internal position estimate and an occasional encoder synchronisation. Because acquisition dominates, the closed-loop iteration time is bounded by the frame period, yielding focus updates at 60 Hz during typical conjunctival imaging. Relative to the prototype before this thesis, the acquisition stage is 3.6× faster (60.2 to 16.67 ms), the computation stage is 20.5× faster (27.5 to 1.34 ms), and the overall focus-adjust cadence improves from 197.0 ms to 16.67 ms, approximately 11.8× faster. Crucially, this improvement is achieved while simultaneously running registration, rendering, and recording threads, none of which interfere with the autofocus pipeline.

By enabling parallel execution of processes, concurrency advances two of the three design goals in this thesis. First, it is the enabling mechanism for design goal 1 (speed): the autofocus loop sustains 60 Hz closed-loop updates because acquisition, computation, and actuation no longer serialise. Second, it enables design goal 3 (clinical accessibility) by allowing frame rendering, recording, registration, and a user interface to operate concurrently with the autofocus algorithm without compromising responsiveness. The remaining lever to improve both speed and autofocus accuracy is the computation stage itself. Therefore,

	Pre-thesis prototype	Current prototype
Acquisition (ms)	60.2	16.67
Computation (ms)	27.5	1.34
Actuation (ms, for 1cm)	231.5	116.0
Overall (ms)	197.0	16.67
Overall (Hz)	(5.1)	(60)

Table 4.1 Measured timings of the autofocus pipeline before and after this work. “Actuation” is the recorded time to achieve a 1 cm translation during a test of ±1 cm moves. “Overall” is the mean interval between focus adjustments observed during a conjunctival angiogram.

the next section focuses on the LoBF algorithm, where algorithmic choices and numerical details produce an accurate LoBF calculation in 1.34 ms.

4.3 LoBF algorithm

The LoBF algorithm is summarised visually in Figure 4.3. Given a frame from the tilted sensor, it returns the horizontal location on the sensor at which the corresponding vertical line of pixels is sharpest. The figure also annotates the mathematical operations linking each stage, and the frame-by-frame computational cost is displayed in Table 4.2. Timings were computed with startup/shutdown outliers removed using an interquartile-range method; averages that include those outliers are approximately ten percent higher. The total frame-to-LoBF latency is 1.34 ms, which corresponds to a processing rate of roughly 746 Hz.

The algorithm proceeds as follows. First, *pre-processing* resizes the frame by half, truncates saturated pixel intensities, applies contrast-limited adaptive histogram equalization (CLAHE, with clip limit 2.0, grid 4×4), and filters with a Gaussian blur (3×3, $\sigma = 1$). Taken together, these steps account for about 29 percent of the runtime (388.0 μs; see Table 4.2) and are detailed in the next subsection. Second, convolution with Roberts Cross operators produces a *sharpness image S*, where pixel intensity depends on gradient sharpness. This is the dominant computational step at 49 percent of runtime (658.8 μs; Table 4.2). Third, a sliding window generates 16-pixel-wide vertical strips centred at every horizontal pixel (except for the edges where this isn’t possible), and a Hann-windowed average converts these strips into a per-column sharpness score. The resulting curve of sharpness versus horizontal position (the orange line in Figure 4) accounts for about 11 percent of runtime (142.9 μs; Table 4.2) and is discussed in Section 4.3.3 on computing the sharpness curve. Here, if a frame has insufficient sharpness anywhere in the image, it is discarded. Finally, the sharpness curve is reduced to a single LoBF value in a manner that yields both accurate estimates and

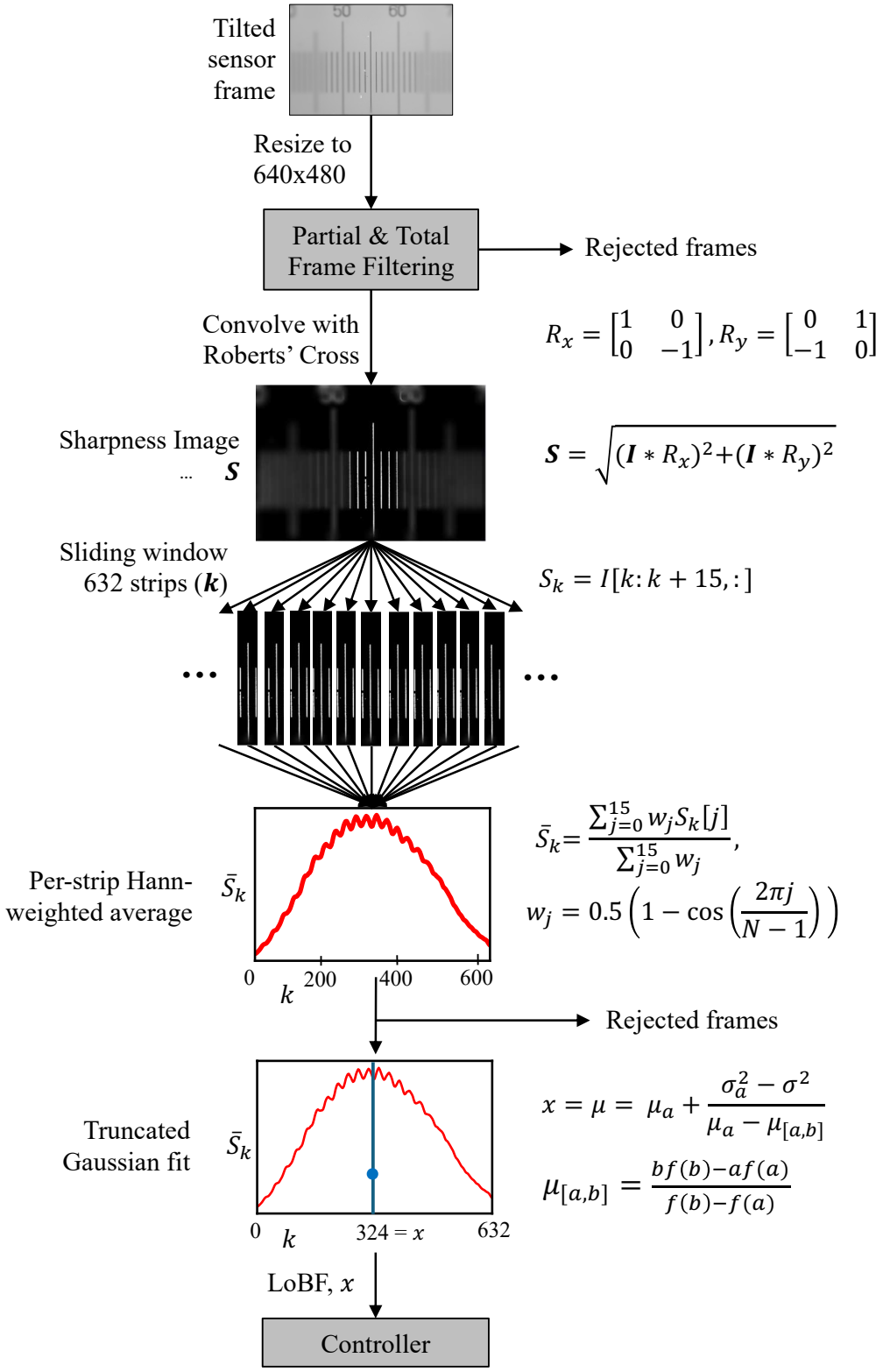


Fig. 4.3 Visual representation of how the LoBF is determined as a position along the horizontal of the tilted sensor.

smooth temporal evolution, avoiding unstable jumps. This last step is negligible in cost (0.1 percent; 2.02 μs) but significantly improves control stability and accuracy, and it is described further in Section 4.3.3.

Table 4.2 Per-frame runtime for each stage of the LoBF algorithm. Timings use interquartile-range outlier rejection; averages including outliers are approximately 10% higher.

Stage	**Runtime (μs)**	**Share (%)**
Resize	110.5	8.2
Pre-process	277.5	20.7
Generate sharpness image	658.8	49.0
Generate sharpness curve	142.9	10.6
Reduce to a LoBF	2.02	0.1
Total	**1343.4 μs (1.34 ms)**	**100.0**

This pipeline is primarily engineered for speed because minimizing sensing-to-actuation latency is essential for stable, high-bandwidth control. While prioritising low latency, it also resolves the accuracy and robustness limitations identified in the prototype available at the start of this thesis (Section 1.2), namely:

1. The algorithm could not manage specular reflections.
2. The algorithm often failed when one side of the image was brighter than the other.
3. The algorithm depended on well-dispersed edges to determine the location of best focus.
4. The system was too insensitive, failing to move the lens despite a slight loss of focus. It was also occasionally too sensitive, leading to uncontrolled oscillation and instability in the lens actuator.

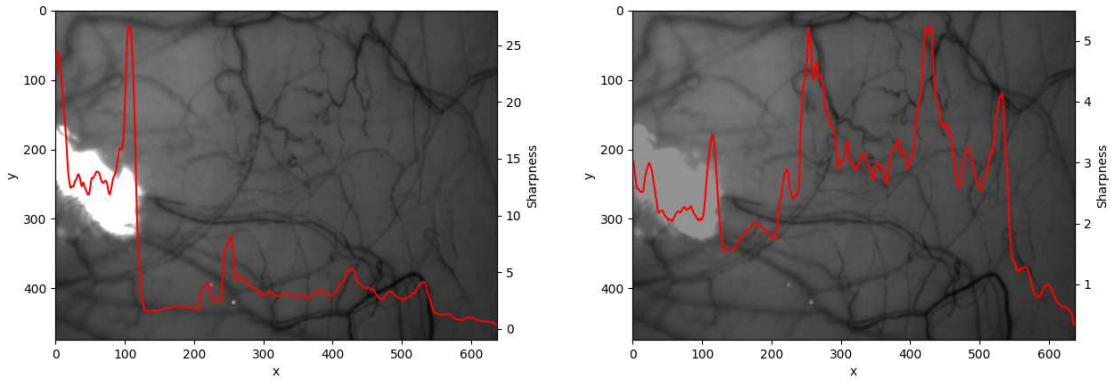
The sections that follow explain the components of the LoBF algorithm that enable fast, robust LoBF estimation. In each section we note the engineering choices that reduce computation time while addressing the accuracy issues above.

4.3.1 Pre-processing

Pre-processing comprises resizing, saturated-pixel truncation, CLAHE (clip limit 2.0, grid 4×4), and Gaussian blurring (3×3, $\sigma = 1$). The degree to which we resize presents a direct trade-off between accuracy and speed. Early prototypes operated at 320×240 (quarter resolution) to approach video-rate, but with a notable loss in edge detail and therefore accuracy. As subsequent engineering improvements accelerated the LoBF pipeline, the analysis

resolution was increased to 640×480 (half resolution), which provided a better balance between computational cost and edge fidelity. At current speeds, full-resolution processing would require approximately 5.36 ms (four times 1.34 ms), which is still compatible with 60 fps input. However, the added latency of full-resolution frames degraded closed-loop stability: the small gain in sharpness precision was offset by the longer response delay to motion-induced artefacts. Consequently, a 50 percent resize is optimal in practice. Resizing itself takes only 110.5 μs on average and is more than recouped by downstream savings in processing time (Table 4.2).

Specular reflections manifest as saturated regions that induce spurious high-contrast boundaries at their edges, confounding sharpness measures. To address this (*issue 1* from the Introduction), an undergraduate project student tested a fast filter that replaces pixels exceeding a chosen intensity threshold with the local mean of surrounding pixels. Figure 4.4a (from [37]) shows a representative specular artefact and an overlaid Tenengrad sharpness measure in red, condensed along the horizontal. The unprocessed image incorrectly peaks at $x = 112$ due to the saturated patch on the left. After truncation (Figure 4.4b), the best-focus location shifts to a more plausible $x = 423$.

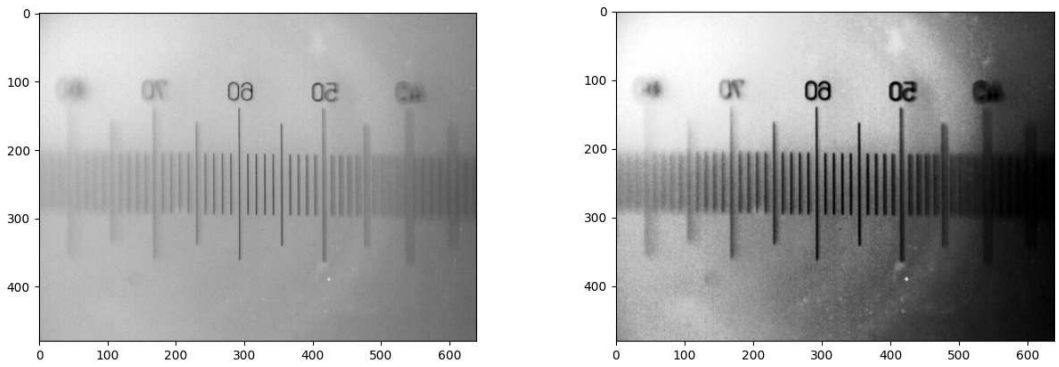


(a) Unprocessed (b) With pixel intensities truncated

Fig. 4.4 Specular reflection rejection technique with the transverse sharpness profile overlaid on in red, from [37].

This simple procedure proved both robust and fast, eliminating the bias and avoiding any LoBF instability attributable to specularities.

To mitigate brightness asymmetry (*issue 2* from the Introduction), a global histogram equalization (HE) step was first evaluated but found unsuitable. When illumination is uneven, HE tends to wash out the bright side and crush the dark side, as illustrated in Figure 4.5.



(a) Without histogram equalization (b) With histogram equalization

Fig. 4.5 Ruler graticule with and without histogram equalization, from [37].

This behaviour follows from the assumption of roughly uniform illumination and its primary intent to stretch global contrast rather than correct spatially varying bias. In contrast, contrast-limited adaptive HE (CLAHE) operates locally and proved effective and robust to global illumination asymmetry. Empirically, a 4×4 grid and clip limit of 2.0 normalised brightness without over-amplifying contrast. Nevertheless, in low-texture regions CLAHE can amplify camera speckle, introducing fine-scale noise. A subsequent Gaussian blur (3×3, $\sigma = 1$) efficiently suppresses single-pixel speckle while preserving edges, whose intensity transitions extend across multiple pixels. Together, pixel-intensity truncation, CLAHE, and Gaussian blurring take on average 277.5 μs and substantially improve robustness for a modest computational overhead (Table 4.2).

With a resized, contrast-enhanced, de-noised, and specular-reflection-free frame, the algorithm proceeds to compute sharpness across the image. Several candidate operators were evaluated for this purpose, as discussed in the next section.

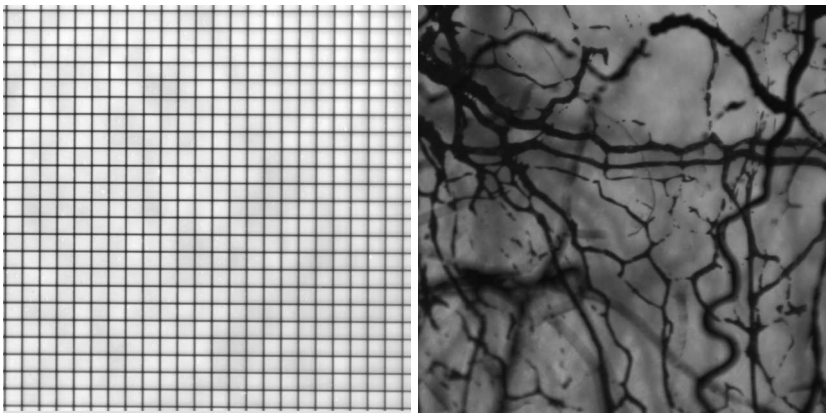
4.3.2 Computing the sharpness curve

Accurately determining the LoBF along the image horizontal requires reducing the image to a sharpness curve that captures how focus varies across the sensor width. We begin by applying a sharpness operator to the input frame to produce a per-pixel 'sharpness image' S. The sharpness curve is then obtained by collapsing S along the vertical dimension. A simple column-wise average was found to be unsuitable because columns containing strong edges dominated, leading to a highly noisy curve with many local maxima. Instead, a local averaging step is required: we consider a neighbourhood of 16 pixel columns for each horizontal pixel and compute a weighted average using a Hann window. This weighted

averaging across nearby columns suppresses noise while preserving focus transitions. The 16-column width was selected empirically as a balance between smoothing and spatial resolution.

Earlier versions of the LoBF algorithm extracted the vertical strips from the original image and then applied the sharpness operator to each strip. We subsequently found that applying the sharpness operator first to the full image and then dividing into strips preserved accuracy while providing a 16× speed-up for that stage of the algorithm. Varying the window width had a relatively lower impact on accuracy or runtime. This is because convolution with the sharpness operator dominates the processing time rather than the weighted averaging.

Regarding the construction of the sharpness image S, Section 2.2.3 of the Literature Review surveys a range of contemporary sharpness measures, from pointwise to frequency-based formulations. Four leading methods were identified and implemented in C++ for evaluation: Roberts Cross, Vollath F_4, Canny, and Tenengrad–Sobel. To perform benchtop comparisons of these operators, the tunable air-lens was locked so that changing the object distance moved the LoBF along the tilted sensor. In effect, the tilted sensor was operated as a range-finder. Two bespoke graticules (Fig. 4.6) were chosen for this comparison: one grid graticule comprising 10 μm lines spaced 100 μm apart, and one multi-layer graticule made up of bitmap traces of conjunctival microvasculature. The former (Fig. 4.6a) is intended to be a best-case scenario for best-focus estimation, and the latter (Fig. 4.6b) is intended to be closest to clinical reality.



(a) Grid graticule (b) Microvasculature graticule

Fig. 4.6 Graticules customised and printed for testing sharpness algorithms.

Each imaging target was mounted on a Thorlabs micrometre and brought into focus with a Zeiss SL120 slit-lamp. For both 20× and 12× magnifications, the target was stepped forward and backward in 0.5 mm increments until the LoBF traversed beyond the bounds

of the tilted sensor. Frames from the tilted camera were saved at each increment, yielding a database with known displacements along the optical axis.

Each best-focus algorithm was then applied to this database, with accuracy and execution speed assessed for each operator. Figure 4.7 illustrates the results for the grid graticule at 20× magnification.

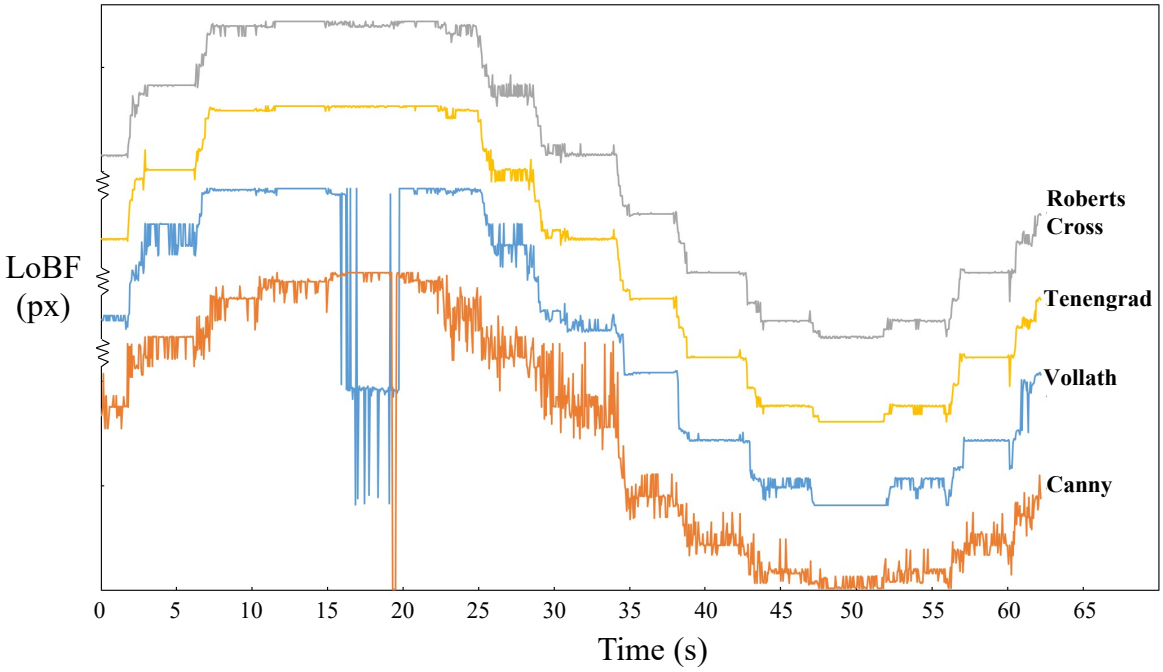


Fig. 4.7 Locations of best focus returned using four sharpness operators as the grid graticule is moved in 0.5 mm increments forward and backward. Traces are combined and offset to facilitate analysis; y-axis is discontinuous.

Tenengrad–Sobel and Roberts Cross produced near-identical outputs, with eleven ±0.5 mm micrometre steps clearly visible. Vollath F_4 also performed acceptably, but yielded incorrect locations as the graticule moved outside the tilted sensor's field. Canny performed poorly. Comparable trends were observed across the other magnifications and graticules.

Table 4.3 summarises the execution times of the respective sharpness operators. These data were obtained using an earlier version of the LoBF algorithm (and thus, do not match the timings in Table 4.1. They are presented here to highlight comparative performance among operators, which is a key metric of interest.

In summary, across Figure 4.7 and other trials, Tenengrad and Roberts Cross were the most accurate, followed by Vollath F_4; Canny was unacceptably inaccurate. From Table 4.3, Vollath F_4 was fastest, followed by Roberts Cross, Tenengrad, and Canny. Because Roberts

Table 4.3 Execution speed of sharpness algorithms on frames of size 320×240 pixels (Intel i5-12600, $n = 4$).

Sharpness algorithm	**Frame rate (Hz)**
Roberts Cross	57.2
Vollath F_4	68.9
Tenengrad–Sobel	43.4
Canny	21.8

Cross matched Tenengrad in accuracy while being substantially faster (57.2 Hz vs. 43.4 Hz, $n = 3$), we selected Roberts Cross for the LoBF pipeline.

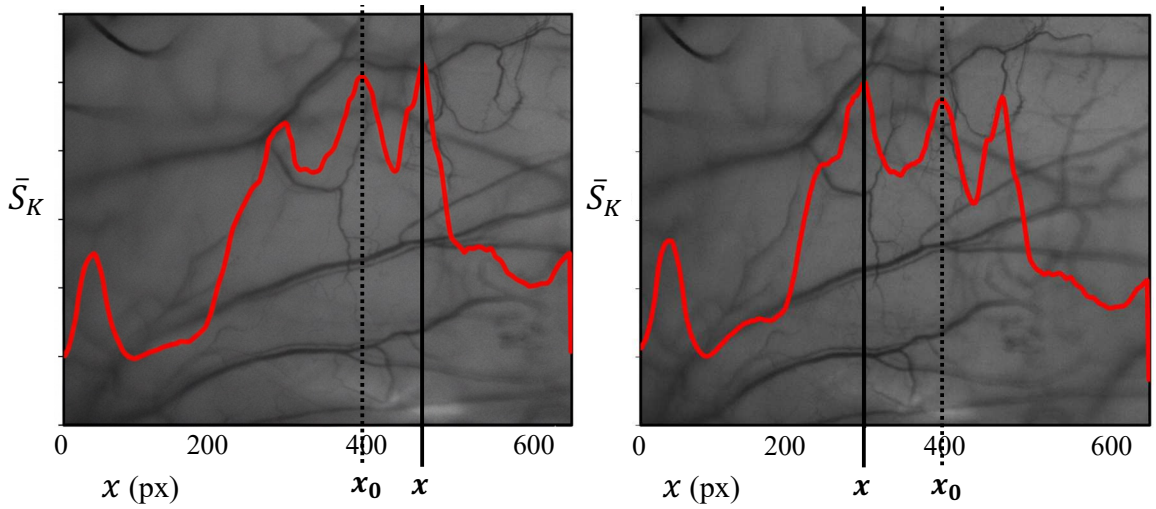
To ensure that subsequent LoBF estimates are based on meaningful image content, a final quality check is applied to each computed sharpness curve. Specifically, the curve's maximum sharpness value must exceed a fixed threshold before reduction to a LoBF proceeds. Frames that fail this criterion are discarded. This simple safeguard proved highly effective in filtering out blink-contaminated or empty frames, since such images contain no region of sufficient contrast to register as in-focus.

4.3.3 Reducing to a LoBF

Early prototypes of the autofocus device defined the LoBF as the maximum of the sharpness curve. While intuitive, this choice contributed directly to the performance degradation observed on sparsely textured targets (*issue 3* from the Introduction). The sharpness metric depends on intensity changes along edges. As a result, the curve is biased toward regions with dense structures, leading to poor estimates for sparsely textured targets.

As the latency of the acquisition–computation–actuation loop was reduced, a second failure mode emerged: the lens frequently oscillated between two nearby locations even though both were slightly out of focus. Figure 4.8 shows two sequential frames (separated by 16.7 ms) from the tilted sensor during a conjunctival angiogram, overlaid with the computed sharpness curve. In both frames the desired LoBF is $x = 320$. Assuming the LoBF is taken as the maximum of the curve, in Fig. 4.8a the LoBF lies just to the left of the desired LoBF, prompting a command to reduce the air–lens separation. One frame later (Fig. 4.8b) the LoBF snaps to the right-hand peak, beyond the target, prompting the opposite command. With local maxima at $x = 263$ and $x = 431$, the LoBF cannot settle at $x = 320$, and the actuator oscillates indefinitely.

More generally, when the sharpness curve contains multiple local maxima, defining the LoBF as the curve's maximum inherently risks oscillation. Several mitigations to this



(a) Frame 1 (b) Frame 2

Fig. 4.8 Subsequent tilted sensor frames from an HVI video, with the sharpness profile $\hat{S}_K$ overlaid, demonstrating why local maxima cause oscillation.

problem were explored. Increasing the vertical sliding-window width from 16 to 32 pixels improved smoothing of the sharpness curve, but local maxima persisted even at the widest window sizes. This occurs because multilayered imaging targets can produce two distinct focus bands on the tilted sensor, yielding two legitimate maxima regardless of smoothing. We therefore sought to reduce the curve to a single representative point that captured changes across the entire sharpness distribution. Such a computation must ensure that the LoBF always varies smoothly in response to small adjustments in air–lens separation.

We first fit a Gaussian to the sharpness curve, as empirical sharpness profiles from the tilted sensor were well-approximated by a Gaussian. For example, the sharpness output for the ideal grid graticule in Fig. 4.6A closely follows a Gaussian with small local maxima at vertical edges. The distribution width σ was pre-determined using an ideal target with well-dispersed edges. The mean μ was then chosen to minimise the squared error between the observed sharpness S_i and the fitted profile $g(i;\mu)$,

$$\sum_{i=0}^{N-1}\left[S_i - A\exp\left(-\frac{(i-\mu)^2}{2\sigma^2}\right)\right]^2,$$

and the estimated LoBF was taken as $x_G = \mu$. Although this enforced a smooth LoBF, it was computationally expensive. Even with fixing the amplitude A to the maximum of S_i, the fit averaged ≈2.1 ms of processing time and scales as $\mathcal{O}(N^2)$. This more than doubles the

LoBF computation time, slowing the response to motion artifacts and increasing controller instability; an unacceptable trade-off.

We therefore adopted a faster centre-of-mass (COM) approach to reduction. A simple COM of the sharpness curve produces a smoothly varying LoBF and accounts for changes anywhere on the curve in $\mathcal{O}(N)$ time. However, despite subtracting an initial bias, the estimator became increasingly inaccurate near the sensor edges, where noise pulled the COM toward the centre. An edge correction based on the distribution's spread mitigated but did not eliminate this effect. Formally, for a sharpness curve S_i we chose

$$\mu_{\mathrm{COM}} = \frac{\sum_{i=0}^{N-1} i S_i}{\sum_{i=0}^{N-1} S_i}, \qquad \sigma = \sqrt{\frac{\sum_{i=0}^{N-1} (i-\mu_{\mathrm{COM}})^2 S_i}{\sum_{i=0}^{N-1} S_i}}, \qquad x_C = \mu_{\mathrm{COM}} + \lambda\,\sigma, \quad \lambda \approx 0.1\text{–}0.3.$$

This method was effective at providing stability when the desired LoBF was toward the centre of the tilted sensor, and it required negligible computational cost. However, residual edge effects required truncating the allowable LoBF to $x \in [190, 450]$ to preserve stability.

To overcome these limitations, we proposed a Truncated-Gaussian (TG) method, which assumes the observed sharpness curve S_i samples a Gaussian distribution truncated over the finite sensor range $[a,b]$. For such a truncated curve, the apparent mean μ_a of the observed data differs from the true mean μ of the underlying Gaussian according to:

$$\mu_a = \frac{\int_a^b x f(x)\,dx}{\int_a^b f(x)\,dx}, \qquad f(x) = A \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

where σ is the experimentally pre-calibrated width of the sharpness distribution. By manipulating the TG integrals, one obtains a correction to recover μ:

$$\mu = \mu_a + \frac{\sigma_a^2 - \sigma^2}{\mu_a - \mu_{[a,b]}}, \qquad \mu_{[a,b]} = \frac{b f(b) - a f(a)}{f(b) - f(a)}.$$

Here σ_a^2 is the apparent variance of the truncated curve, and $\mu_{[a,b]}$ encodes boundary asymmetry over $[a,b]$. In practice, $f(a)$ and $f(b)$ are estimated by averaging sharpness values near the image edges; when $f(a) \approx f(b)$, the correction is negligible and $\mu \approx \mu_a$.

The TG method recovers the mean of the underlying Gaussian even when the focus band is partially clipped by the sensor, yielding a single, smoothly varying LoBF without discrete jumps. In contrast to COM reductions, it is robust to edge bias, so the target LoBF can be set across a wider segment of the tilted sensor. In contrast to least-squares gaussian fitting, the computation cost is negligible: 2.02 μs per frame or $\approx 0.1\%$ of the pipeline, preserving the latency required for high-bandwidth control.

Reducing the sharpness curve to a robust, smoothly moving LoBF based on the spatial distribution of sharpness significantly improves the algorithm's overall accuracy. In practice, this computation step significantly reduces the algorithm's dependence on well-dispersed edges (*issue 3*), as it can account for features and changes across the full sharpness profile. It also eliminates the oscillatory failure mode illustrated in Fig. 4.8. The next section therefore turns to controller design, using the TG-based LoBF to define the tracking error and to drive the difference between the desired and measured LoBF to zero.

4.4 Designing the controller

From the previous section, we obtain an accurate and consistent LoBF in approximately 1.34 ms from each tilted-sensor frame. However, accuracy issue 4 from earlier prototypes (Section 1.2) remained: the system was both insensitive to minor de-focusing and overly sensitive to large movement artifacts. These behaviours, along with other empirical observations, reveal that autofocus at video-rate involves a control problem.

To start, the root cause of the insensitivity was the 100-pixel tolerance band used in the prior zone-based controller. Within this band the image was judged to be in-focus, and no lens commands were sent. In an optimal design, the tolerance band should be just wide enough to absorb the combined imprecision of the tilted sensor's measurement noise and the best-focus algorithm's variance. To quantify this imprecision, the dual-plane graticule in Figure 4.6B was held stationary while the best-focus algorithm ran in a loop with no lens motion. Across five 60-second trials, the measured LoBF varied by up to 6 pixels despite the static scene. This implied a target tolerance band near 6 pixels (as tight as practical). Achieving such a narrow band required replacing the legacy zoning controller.

Next, our fundamental objective is to drive the focus error $e(t)$ to zero, where $e(t)$ is the difference between the desired LoBF and the measured LoBF. Although one could attempt a purely model-based approach by computing a fixed constant from the formulae in Section 3.2, this model necessarily relies on assumptions (e.g., thin-lens, parallel-beam, negligible unmodeled dynamics) that require empirical correction in practice. Those empirical corrections effectively reduce to proportional-gain tuning. We therefore cast the problem explicitly as feedback control, and adopt a classical PID-family architecture. PID-style controllers are computationally lightweight for video-rate operation, robust to moderate plant uncertainty and disturbances, easy to interpret and tune, and are well matched to the actuator–optics dynamics [20]. The resulting control structure is shown in Figure 4.9.

Figure 4.9 depicts the autofocus loop as a standard cascaded architecture adapted from [20]. The outer loop uses a PD controller to compute a relative move command

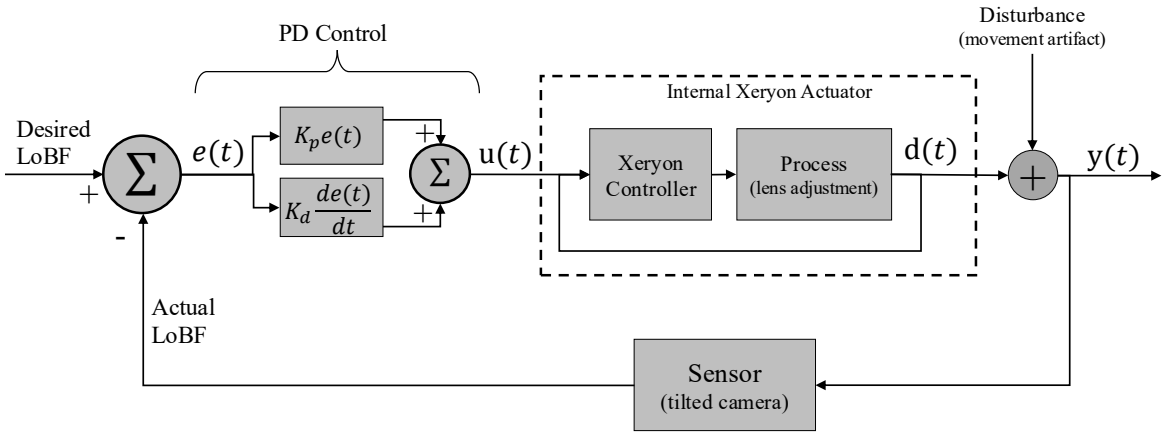


Fig. 4.9 Closed-loop control block diagram for the autofocus device.

$u(t)$ from the focus error $e(t)$, defined as the difference between the desired and measured LoBF. The command $u(t)$ is forwarded to the Xeryon actuator's embedded controller, which is itself a gain-scheduled PID running at tens of kilohertz. This internal loop regulates the physical position of the tunable air-lens and outputs the achieved displacement. Together, the actuator's internal controller and the lens dynamics constitute the *process* block. A disturbance term $d(t)$ enters at the plant output to represent specimen motion or objective motion that perturbs focus independently of the actuator. The resulting image $y(t)$ therefore reflects both the commanded lens motion and any disturbance. The tilted sensor measures the actual LoBF from $y(t)$, closing the outer feedback loop. In this cascade, the high-bandwidth Xeryon loop ensures precise actuator positioning while the outer PD loop accommodates for optical defocus at video rate (60 Hz).

Briefly, the proportional term in a PID controller provides the primary corrective action by producing a control signal proportional to the instantaneous error. The integral term reduces steady-state bias by accumulating error over time, but can degrade stability (increase tendency to oscillate) if misused [20]. The derivative term anticipates error trends and dampens overshoot, which permits higher proportional gains, although it can amplify measurement noise [20]. Consistent with these trade-offs, the integral gain was set to zero ($K_i = 0$). Steady-state pixel-level bias is not a practical concern for this autofocus system because the tolerance band is intentionally wider than residual infinitesimal error. The potential benefit of eliminating residual pixel-level bias is outweighed by the heightened risk of oscillatory behaviour.

By contrast, derivative action is desirable because overshoot and oscillation have been recurring failure modes. While derivative control can be sensitive to noise, the measurement pipeline and the practical tolerance band (implemented at $\sim$6–8 pixels) attenuate much of

that sensitivity. For these reasons, a PD controller in the outer loop best meets the competing requirements of responsiveness and robustness.

During early tuning (next section), we observed that any single proportional gain K_p made the controller either too aggressive for large moves (leading to instability) or too sluggish for small moves. To reconcile these behaviours, we implemented a simple gain-scheduling strategy within ± 100 px about the desired LoBF: K_p is scaled from 100% at the edge of this region down to 30% near the setpoint. This schedule reflects the nonlinear sensitivity of the optical response, as observed in the empirical calibrations from Fig 3.4. Far from focus, sharpness varies slowly with defocus and requires stronger corrective action, whereas near focus, the same gain can induce overshoot due to higher sensitivity and measurement noise. Reducing K_p near the setpoint therefore improves damping and prevents oscillation, while higher gains far from focus accelerate convergence. Qualitatively, this improved stability and responsiveness and mirrors the internal gain scheduling in the Xeryon controller.

In summary, the controller design strives to drive $e(t)$ to zero as quickly as possible while maintaining stability across objectives and transient disturbances. Although a nominal gain could be computed from the analytical model in Section 3.2 (and a measured tube-lens focal distance), the inevitable empirical refinements amount to tuning a PD controller. We therefore proceed with an empirical tuning procedure, described next.

In summary, the autofocus controller comprises a cascaded architecture with a high-bandwidth inner position loop (Xeryon PID) and an outer PD loop that corrects optical defocus in real time. The outer loop's role is to reject focus disturbances while maintaining responsiveness across imaging conditions. Because analytical models cannot fully capture the coupled actuator–optics dynamics, the effective performance depends primarily on empirical selection of the proportional and derivative gains. The next section details the tuning procedure used to determine these parameters.

4.4.1 Tuning the controller

The gain-scheduled PD controller was tuned empirically using a standard Ziegler–Nichols style procedure [124]: K_p was increased until the onset of oscillation, and a derivative term was then introduced to achieve a fast, non-oscillatory transient. Three qualitative behaviours were consistently observed as K_p was varied:

1. With K_p too small, the lens moves sluggishly and the LoBF approaches the centre of the tilted sensor asymptotically.

2. With K_p too large, the lens and the LoBF oscillate before settling; beyond a threshold, the loop becomes unstable and oscillates indefinitely.

3. The optimal K_p depends on the disturbance magnitude (i.e., the graticule displacement): for a given displacement, an optimal K_p achieves the minimum-time return of the LoBF to the sensor centre with no oscillation.

Because the ideal PD gains depend on the size of the focus error, we first characterised typical Z-axis subject movements (reported in section 6.2.1). A Fourier transform of those data revealed no dominant frequencies, indicating that the relevant disturbances are primarily transient patient motions. The largest observed motion was 0.8 mm over 1.2 s. Accordingly, we adopted a conservative tuning target based on a 1 mm transient, which provides a margin of safety. The tuning objective was to minimise the response time to this worst-case transient while avoiding overshoot. Under this strategy, the autofocus mechanism is guaranteed not to oscillate for near-instantaneous patient motions of 1 mm or less.

To approximate a 1 mm step disturbance reproducibly, the grid graticule was mounted on a spring-loaded slide held in tension by a 1 mm-thick pin. Removing the pin caused the graticule to snap 1 mm away from the slit lamp, after which the autofocus loop responded. The snap motion occurred over approximately 30 ms ($n = 3$), i.e., about two frames at 60 Hz, which closely approximates a step input. For each run, the LoBF on the tilted sensor was recorded; representative traces are shown in Figures 4.10 and 4.11.

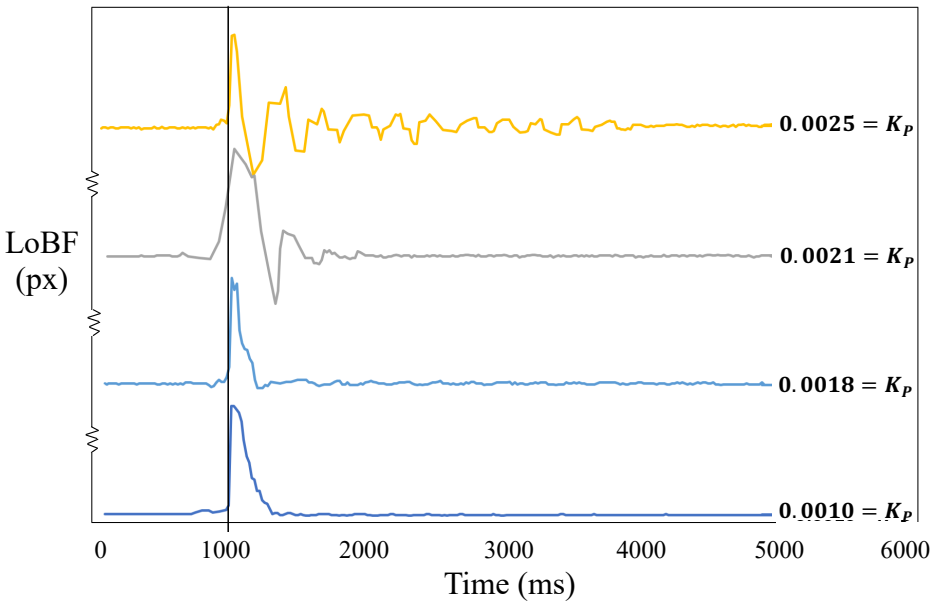


Fig. 4.10 Proportional controller response for a variety of K_p values. Vertical line marks 1mm disturbance onset.

Using a P controller alone ($K_d = 0$) and a 1 mm near-step, K_p could be raised to ≈ 0.0018 while still avoiding overshoot (Fig. 4.10). With derivative action enabled, K_p could be

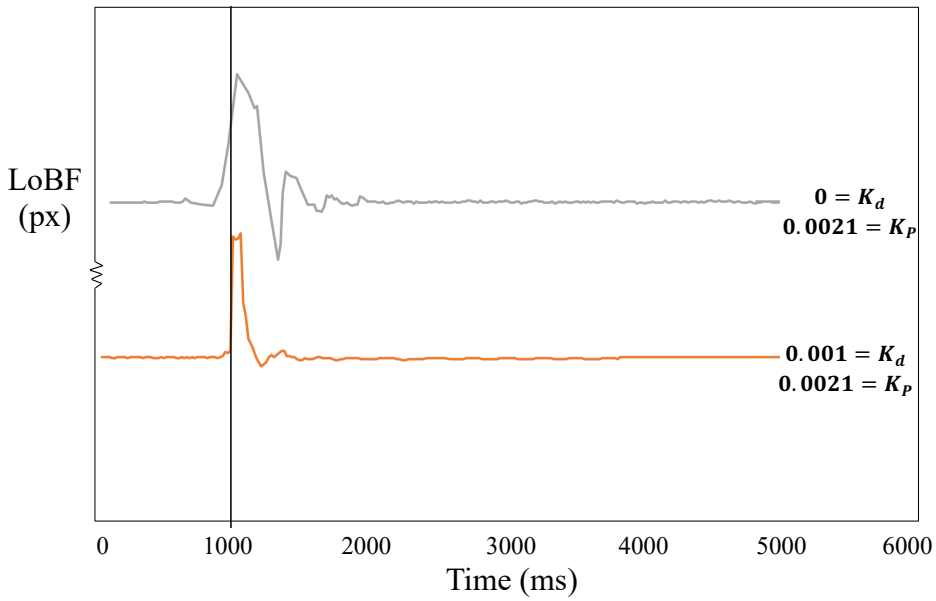


Fig. 4.11 Proportional-derivative controller response to a 1mm disturbance for a variety of K_p, K_d values. Vertical line marks 1mm disturbance onset.

increased to 0.0021 paired with $K_d = 0.001$ (Fig. 4.11). Using the conservative settling criterion of remaining within ± 15 px of the desired LoBF, the loop settles in $\approx t_s = 142$ ms (n=3) for a 1 mm near-step disturbance. The P-only controller is stable but exhibits a longer settling time and softer transients. On this basis, we selected $K_p = 0.0021$ and $K_d = 0.001$ for the autofocus controller. These gains are intentionally somewhat aggressive, but stability near the setpoint is preserved by the gain-scheduling strategy described in the previous section.

As shown in Section 4.6, the K_p parameter is exposed to the user through the *Settings* dialogue in case minor in-field adjustment is necessary. Such adjustments may be needed for changes in the tube lens (e.g., when switching microscope platforms). It also may be necessary when compensating for orientation-dependent gravitational effects in the Xeryon actuator, which in some mounting orientations benefits from a slightly larger K_p. These adjustments are discussed as limitations in Chapter 7.

4.5 Real-time image registration

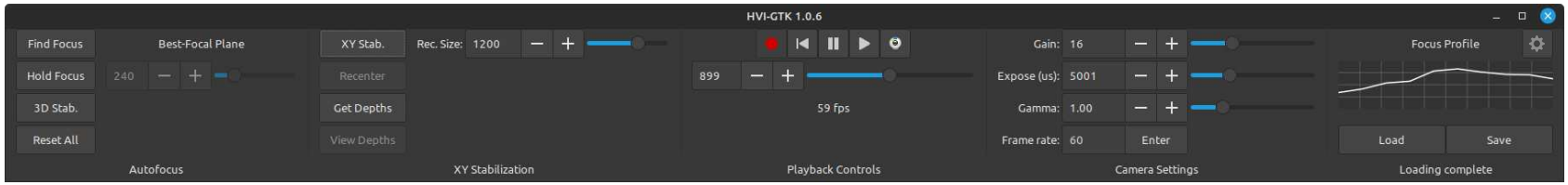
To deliver stable video microscopy in three dimensions, the system must counter both axial defocus and in-plane motion. Executing Z-axis autofocus together with X–Y alignment on frames streaming at 60 Hz is non-trivial. Despite the computing headroom provided by the hardware from Section 3.6, the efficient LoBF algorithm provided by Section 4.3, and the multi-threaded code-base provided by Section 4.2, the existing X–Y stabilization pipeline

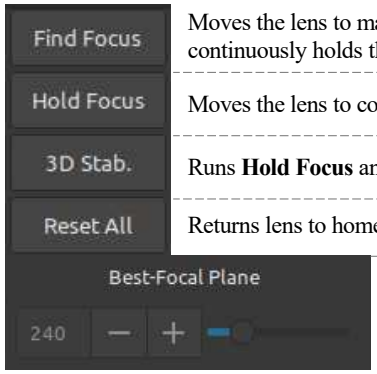
exceeded the frame-time budget. We thus sought a registration approach compatible with concurrent autofocus.

Earlier prototypes used a bespoke stabilization method developed by Tom Drummond for image registration, which models X-Y motion as a robust, global translation estimated from a sparse set of vessel-like edges. Each incoming frame is Gaussian-smoothed at a user-selected scale and analysed via the image Hessian; the largest eigenvalue provides a ridge strength map while its eigenvector gives principal orientation. Non-maximum suppression along that orientation yields sub-pixel edges that are linked into chains, and a random subset of edges is then tracked frame-to-frame.

In our system this method achieved real-time operation with the Prosilica GC1380H, which streams 1.4 megapixel frames at 30 fps. However, it required user input to select scale and threshold parameters before running, and it entailed a sizable per-frame computation cost. With the newer Alvium 1800 U-291m, frame size increased to 2.9 MP and frame rate increased to 60 fps, increasing pixel throughput and cadence by around 4.1 times. Drummond's pipeline no longer met the frame budget, and its reliance on content-dependent parameters (threshold, scale) made behaviour brittle across scenes and gain/exposure settings. These considerations motivated a shift to a frequency-domain approach with lower computational cost and fully automatic operation.

We therefore adopted a phase correlation technique for real-time image registration following the literature review in Section 2.3. Each frame is Hann-windowed and transformed to the frequency-domain, where the cross-power spectrum with a cached reference FFT is normalised to unit magnitude, inverse-transformed to a correlation surface, and the wrapped peak gives the integer shift. A quadratic 3x3 fit around this peak provides sub-pixel refinement, and an exponential moving average updates the reference to track slow drift. This keeps the per-frame computation to one forward FFT on the current frame and one inverse FFT on the cross-power spectrum. To meet the 60 fps budget we combined several additional techniques: spatial down-sampling (registering at half-resolution and applying the shift at full resolution), temporal decimation (estimating the shift on every other frame and linearly interpolating offsets for intervening frames), and caching all static terms (reference FFT and window). Together these changes removed the need for manual tuning, and delivered robust, real-time registration at the higher resolution and frame rate. Crucially, it is computationally inexpensive enough to operate in parallel with the autofocus and GUI processes at 60 fps, allowing Z-axis autofocus synchronised with X-Y registration at video rate.





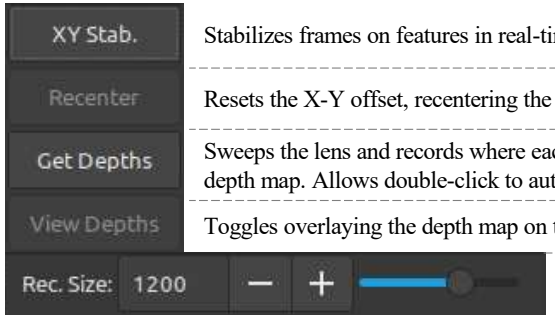
Moves the lens to maximize sharpness across the image, and then continuously holds that focal plane.

Moves the lens to continuously hold the *current* focal plane.

Runs **Hold Focus** and **XY Stab.** simultaneously.

Returns lens to home position, and resets any frame offset.

Becomes active when **Hold Focus** or **Find Focus** are active.
Sliding to left shifts the focal plane deeper, and to the right shifts it shallower.



Stabilizes frames on features in real-time by applying an X-Y offset.

Resets the X-Y offset, recentering the frames.

Sweeps the lens and records where each feature is sharpest, creating a depth map. Allows double-click to autofocus on a specific feature.

Toggles overlaying the depth map on the scene.

Recording size in # of frames.

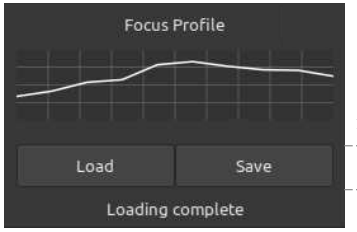
Starts or continues a recording.

When rewatching a recording, returns to the start.

Pauses a recording or pauses playback of a past recording.

Plays playback of a recording.

Returns to the live camera feed from a loaded recording.



Shows a sharpness profile across the image horizontal, used to ensure the microscope is orthogonal.

For loading and saving recordings.

Load/Save progress indicator.

Provides user access to various settings, including:
1. **Controller gain**: Lower for a slower, more stable response.
2. **Find Focus** home: the location on the tilted sensor that corresponds with a sharp imaging camera, typically near the center. Used when finding focus.

Fig. 4.12 Annotated overview of the autofocus software graphical user interface (GUI).

4.6 User interface and operation

An accessible user interface was crucial for achieving our third design goal: clinical accessibility. The UX is made up of one large window (*SDL Window* in Figure 4.1) for image streaming, and a smaller GUI window with buttons and playback features (*GTK Window* in Figure 4.1). The larger image-streaming window supports mouse-wheel zooming, click-and-drag panning, and dynamic resizing, enabling intuitive navigation during live imaging. It is programmatically linked to the smaller GUI window such that both open and close as a single, unified application instance. This is a common application layout for image-processing software like ImageJ and Fiji.

The autofocus workflow centres on two modes. *Find Focus* moves the lens to maximise sharpness across the image and then continues to hold that focal plane by setting the desired LoBF to where the imaging camera is maximally sharp (typically near the centre of the tilted sensor). *Hold Focus* holds the current focal plane by using the current LoBF as the setpoint, allowing an operator to lock onto a chosen structure even when it does not dominate the scene. In either mode the *Best-Focal Plane* slider becomes active; dragging left shifts the setpoint deeper and dragging right moves it shallower. This slider enables deliberate, precise offsets from the moving nominal plane. *Reset All* returns the refocusing lens to home and clears any accumulated frame offsets.

Planar stabilisation is available on demand. *XY Stab.* applies the phase-correlation registration described in Section 4.5 to render position, removing lateral motion in real time. *Recenter* zeros the accumulated offset. *3D Stab.* runs *Hold Focus* and *XY Stab.* concurrently, maintaining axial focus while stabilising in X–Y at video rate. This allows three-dimensional motion during acquisition to be automatically managed.

Recording and playback are integrated as controls across the top bar: record, rewind-to-start, pause, play, recording size, and return-to-live interactive elements provide all that is necessary. Files can be *Load*ed and *Save*d to the hard drive, with a progress indicator for longer transfers. This enables repeatable protocols in which identical sequences are captured, annotated, and re-examined without leaving the acquisition environment.

Crucial camera configuration settings are surfaced: the *Gain*, *Exposure (µs)*, *Gamma*, and *Frame rate* controls expose parameters used most frequently for digital microscopy. A compact *Focus Profile* plot shows sharpness across the image horizontal, which operators use to confirm that the microscope is orthogonal to the target before beginning a sequence. Advanced settings are gathered under the gear icon. For example, the *Controller gain* parameter exposes K_p for the outer PD loop, allowing minor adjustments as discussed in Section 4.4.1.

Finally, the platform includes a depth-aware feature to improve feature-based autofocus. Activating *Get Depths* sweeps the refocusing lens and records, for each prominent feature, the lens position at which it is sharpest, producing a depth map that can be overlaid with *View Depths*. After this calibration, a double-click anywhere in the live image instantly locks focus onto that feature, even as it moves, by driving the LoBF to the stored depth. This is the first operational bridge between the imaging camera and the tilted sensor and foreshadows sensor-fusion strategies discussed in Section 7.3.2 as an opportunity for future development. A full operating manual is provided in Appendix A and has been supplied to operators at Addenbrooke's for clinical use.

4.7 Conclusions

This chapter has presented the software layer that turns the hardware of Chapter 3 into a reliable, video-rate, image-based autofocus instrument. We re-architected the acquisition–computation–actuation loop for concurrency, replaced prototype scripts with a C++ codebase, and implemented a robust LoBF pipeline coupled to a gain-scheduled PD controller. Alongside autofocus, we added real-time phase-correlation image registration, initial sensor-fusion hooks, and a streamlined desktop application. Collectively, these elements close the sensing-to-actuation loop at video rate while preserving GUI responsiveness, recording, and image registration. Ultimately, this chapter moves the system from a proof-of-concept to a deployable tool.

With respect to the first design goal of *speed*, the design meets the target of sustained 60 Hz closed-loop operation. The concurrent threading model decouples acquisition, LoBF computation, and actuation, eliminating the blocking that constrained earlier prototypes. Within this architecture the LoBF pipeline executes in 1.34 ms end-to-end, leaving the iteration time bounded by the 16.67 ms frame period. Relative to the pre-thesis prototype, acquisition is 3.6× faster, computation is 20.5× faster, and the effective focus-adjust cadence improves by 11.8×, and these gains are realised while registration, rendering, and recording run in parallel.

Regarding the second design goal of *accuracy and robustness*, algorithmic choices in the LoBF pipeline address the specific failure modes identified at the outset. Pre-processing removes specular artefacts and corrects illumination asymmetry; the Roberts Cross operator provides an accurate, high-throughput sharpness image; and a quality gate rejects frames lacking sufficient texture. Critically, a truncated-gaussian approach reduces the sharpness curve to a LoBF that moves smoothly and remains unbiased near sensor edges, eliminating the oscillatory behaviour that arose when using a raw maximum. The outer-loop controller

provides fast convergence without overshoot to worst-case transient disturbances. Together, these choices deliver precise, stable tracking across objectives and scenes.

The third design goal, *clinical-grade accessibility*, is supported by a well-designed codebase and complete user interface. The desktop application provides two autofocus options for operators, *Find Focus* and *Hold Focus*, paired with an intuitive setpoint slider for deliberate focal plane selection. It also combines axial autofocus with optional automatic X–Y stabilisation to manage three-dimensional motion during routine use. The *Get Depths* workflow establishes a practical bridge to sensor fusion by linking imaging-camera features to the tilted sensor's depth coordinate. In the codebase, bounded queues, timestamps, and conservative scheduling keep the user interface responsive under load. The conventional C++ layout facilitates installation, inspection, and extension by other research teams.

In synthesis, this chapter has shown how a performant software architecture converts capable hardware into a high-bandwidth, clinically usable refocusing platform. The latency budget is now dominated by acquisition rather than computation or control, the LoBF estimate is both rapid and accurate, and the controller is fast and stable. The result is a system that achieves the intended performance envelope without imposing a configuration burden on the operator.

The next step is to quantify the system's performance, and then demonstrate its broad applicability. Chapter 5 quantifies the device's performance under controlled conditions, examining the settling time to varying disturbances, optical effects on image quality, and the operational range across objectives. Chapter 6 translates this characterisation into practice. Demonstrations include conjunctival imaging on two different slit-lamps, benchtop microscopy of deforming materials, and focus-holding across a variety of imaging targets.

Chapter 5

Device performance characterization

5.1 Introduction

The preceding Chapters 3 and 4 described the hardware and software that realise a universal, video-rate, image-based autofocus system. Building on that foundation, this chapter characterises the device in benchtop experiments designed to verify that the implemented architecture performs as intended in practice. The emphasis is on system-level behaviour rather than component-level minutiae, with experiments chosen to probe the three design requirements established earlier: speed, accuracy/robustness, and clinical applicability. Guided by these aims (and constraints on experimental time), we focus on four questions. First, does the closed-loop controller reject transient axial perturbations rapidly and accurately? Second, how does inserting the autofocus device into the optical path affect image quality–particularly at the level of individual erythrocytes? Third, does the device preserve magnification as the air-lens separation changes? Fourth, what is the practical working range across microscopes and objectives? Each question is addressed with targeted benchtop tests that isolate the relevant behaviour while remaining representative of clinical and laboratory use.

Section 5.2 quantifies closed-loop settling dynamics by imposing controlled step-disturbances and measuring the response. Section 5.3 then compares image formation with the autofocus device inserted upstream of the imaging camera against a direct camera-only baseline. Using this experimental setup, Section 5.3.1 evaluates resolution and sharpness using a $10\,\mu$m dot target, while Section 5.3.2 assesses the effective depth of field by translating the same target in 0.05 mm increments toward and away from best focus. Section 5.4 examines magnification constancy to confirm that image scale remains effectively invariant as the air-lens traverses the full focusing range. Finally, Section 5.5 characterises the operational range of the device across three microscope platforms and eleven objectives in total, demonstrating reliable

performance across diverse optomechanical contexts. Together, the results provide a concise experimental validation of the autofocus system's speed, fidelity, and general applicability.

5.2 Controller settling times

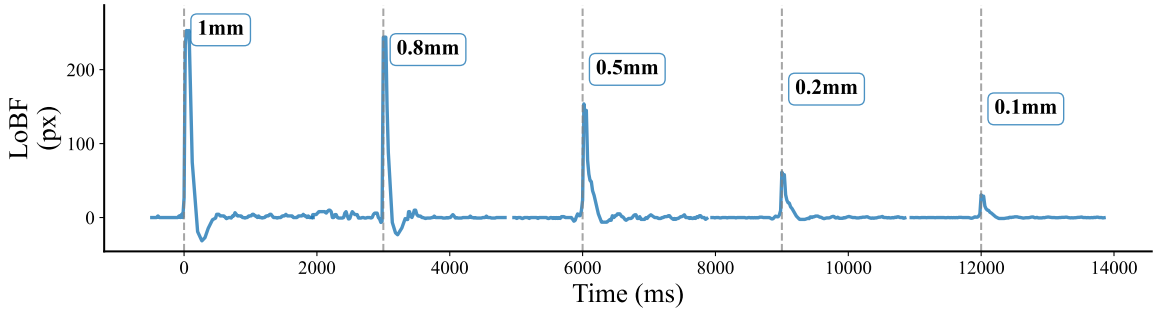


Fig. 5.1 LoBF stream from the tilted sensor following near-step object-space disturbances with active air-lens accommodation. Vertical dashed lines mark disturbance onsets; callouts indicate displacement magnitude.

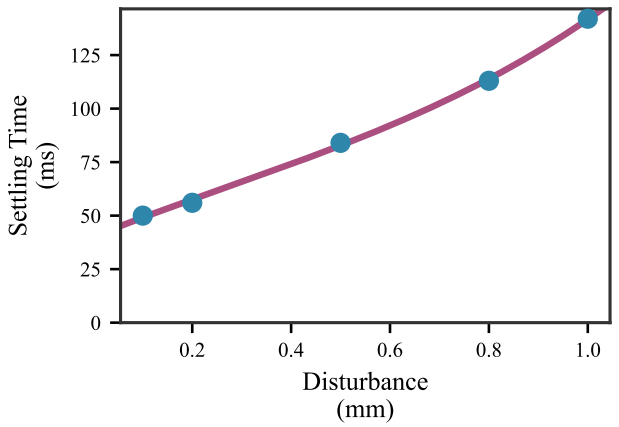


Fig. 5.2 Settling time t_s: time required for the LoBF to return and remain within ± 15 pixels following a near-step disturbance.

In Section 6.2.1 we characterised patient motion during conjunctival angiography. Across eight eyes from four participants, the largest transient was $\approx 0.8\,\text{mm}$ over $1.2\,\text{s}$, and spectral analysis revealed no dominant noise components above $1\,\text{Hz}$. Therefore, most artifacts arise from slow drifts or isolated transients, making step-like perturbations an appropriate probe of closed-loop behaviour.

Therefore, we quantified the closed-loop dynamics by imposing known object-space displacements using a spring-loaded stage and measuring the LoBF from the tilted sensor. Spacers of $\Delta d = \{0.1, 0.2, 0.5, 0.8, 1.0\}$ mm were removed to create near-step inputs (Fig. 5.1). With a 60 Hz acquisition (camera-limited), $\Delta d = 0.1-0.5$ mm unfolded within one frame (≈ 16 ms), while $\Delta d = 0.8-1.0$ mm spanned two frames (≈ 32 ms), both effectively mimicking step disturbances. The stream of LoBF's from the tilted sensor in response to these transients are plotted on one domain in Figure 5.1. We repeated the experiment with image-space step inputs (rather than object-space) by digitally shifting the LoBF set-point by the corresponding number of pixels. Results were similar, with slightly faster settling at $\Delta d = 0.8$–1.0 mm due to the absence of frame lag. Only object-space data are reported here, as they best reflect clinical use.

The proportional–derivative (PD) controller (Section 4.4), tuned using a Ziegler–Nichols procedure, returned the system to its focus set-point without detectable steady-state error on all trials. It also maintained the optical focal plane indefinitely with no evidence of thermal drift. Steady-state LoBF jitter never exceeded the sensor noise (± 6 px). Defining settling time as the interval until the LoBF re-entered and remained within a ± 15-pixel band around steady state, the controller achieved t_s from 50 ms at $\Delta d = 0.1$ mm up to 142 ms at $\Delta d = 1.0$ mm (Fig. 5.2). Small overshoots occurred only for $\Delta d = 0.8$ and 1.0 mm and were $< 25\%$ of peak. Settling time increased superlinearly with disturbance magnitude, consistent with modest overshoot and air-lens inertia. However, attempts to reduce overshoot further came at the cost of longer t_s overall, so minor overshoot is acceptable for these worst-case transients. Gain-scheduling improved high-disturbance settling while preserving stability across objectives and microscope platforms.

5.3 Effect on image quality

A central concern with the proposed autofocus system is whether it compromises image beam quality, which would be unacceptable in high-quality digital microscopy. Two mechanisms could, in principle, degrade imaging performance: (i) $\sim 10\%$ of the photons are intentionally directed toward the tilted sensor and thus away from the imaging camera, potentially reducing signal-to-noise ratio; and (ii) additional optics could introduce aberrations that broaden or distort the point spread function. To quantify any such effects, we required a camera-only control view that would allow image-for-image comparison against the autofocus (AF) path.

All imaging in this section was performed on the Zeiss SL120 slit-lamp, which provides nominally identical dual camera ports, enabling a side-by-side arrangement. We mounted a standard Zeiss C-mount adaptor and the same imaging camera (Alvium 1800 U-291) on

the opposite port from the AF device. We used 20× magnification, as this is the clinically relevant magnification used throughout the experiments in Section 6.2. Illumination was oriented coaxially, to avoid any parallax or shadow on the imaged graticule. Across all experiments we held acquisition parameters fixed (gain = 14, exposure = 6 ms, gamma = 1.00, frame rate = 60 fps). Images from the camera-only and camera-with-autofocus paths were then collected and analysed.

This comparison is not perfectly symmetric. The Zeiss C-mount arm exhibited a 39.9% higher magnification at the imaging camera and a correspondingly lower image luminance, despite identical camera settings. These differences were explicitly accounted for prior to quantitative analysis: images were first rescaled to a common magnification (i.e., the higher-magnification view was down-sampled by a factor of $1/1.399$ so that object-space dimensions matched), and intensities were normalised by a single global multiplicative factor so that background levels were comparable across views. Because the primary sharpness metric used below—the Gaussian-fit full width at half maximum (FWHM)—depends on the spatial spread rather than absolute intensity, this normalization does not affect the resulting FWHM estimates. With these controls in place, we proceed to assess resolution.

5.3.1 Resolution

To test whether the autofocus optics compromise image formation at best focus, we imaged a 10 μm dot graticule with and without the autofocus module in the optical path. A 10 μm target was chosen to approximate erythrocyte dimensions [19]. Representative frames are visually similar between the two paths (Fig. 5.3).

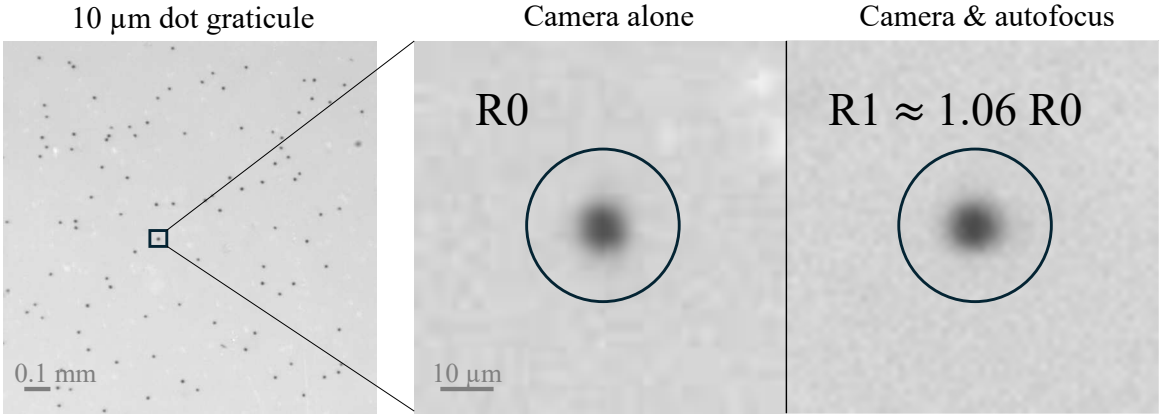


Fig. 5.3 A bespoke 10 μm dot graticule (left) cropped and enlarged for the camera alone (middle) and for the camera with the autofocus device upstream (right).

Quantitatively, two-dimensional Gaussian models were fit to the dot images, and full-width at half-maximum (FWHM) values were extracted along horizontal and vertical line profiles. Because the peak sharpness was not strictly centred in either image, we evaluated line profiles over a small neighbourhood around the geometric centre (rows for horizontal width; columns for vertical width) and reported the *maximum* FWHM per axis to provide a conservative, centre-bias–resistant estimate.

For the camera-only path, the maximum horizontal and vertical FWHM were 35.0 and 34.0 pixels, respectively. For the AF path, the corresponding maxima were 26.0 and 26.3 pixels, respectively. To compare in object-space, the camera-only measurements were scaled by the magnification ratio (39.9%), yielding corrected maxima of 25.0 pixels (horizontal) and 24.3 pixels (vertical). The AF-to-camera ratios were therefore

$$\frac{\mathrm{FWHM}^{x}_{\mathrm{AF}}}{\mathrm{FWHM}^{x}_{\mathrm{cam,corr}}} = 1.04 \quad \text{and} \quad \frac{\mathrm{FWHM}^{y}_{\mathrm{AF}}}{\mathrm{FWHM}^{y}_{\mathrm{cam,corr}}} = 1.08,$$

i.e., the camera-only path is approximately 4% sharper horizontally and about 8% sharper vertically (AF broader by $4-8\%$).

From a practical standpoint, an 8% broadening at a $10\,\mu$m feature scale is modest and remains compatible with accurate visualization of erythrocyte-scale structures, especially given that the analysis was performed in object space after correcting for the 39.9% magnification difference. The observed brightness offset between ports was handled by global intensity normalization, which does not alter fitted widths and thus does not bias the FWHM comparison. Taken together, these results show that inserting the AF device upstream of the imaging camera does not materially degrade in-focus spatial resolution for the use case of interest.

To probe for anisotropic broadening (astigmatism), we quantified the eccentricity of the fitted point-spread function for a representative $10\,\mu$m dot. Each dot was modeled with an elliptical two-dimensional Gaussian parameterised by σ_x, σ_y, and the major-axis orientation ϕ. Eccentricity was defined as

$$e \;=\; \sqrt{1-\left(\frac{\sigma_{\min}}{\sigma_{\max}}\right)^{2}}, \qquad \text{with} \quad \mathrm{FWHM} = 2\sqrt{2\ln 2}\,\sigma,$$

so that $e=0$ corresponds to a circular (isotropic) spot. The fitted eccentricities were low and comparable between paths: camera-only $e=0.09$ and AF-inserted $e=0.07$, corresponding to axis-ratios $r=\mathrm{FWHM}_{\max}/\mathrm{FWHM}_{\min}$ of approximately unity for both paths. Major-axis orientations differed by only $\Delta\phi = 0.1^{\circ}$, indicating that the observed $\sim 8\%$ increase in

FWHM arises from isotropic broadening rather than astigmatism. These results confirm that the AF optics introduce no measurable astigmatism or directional blur.

5.3.2 Depth of field

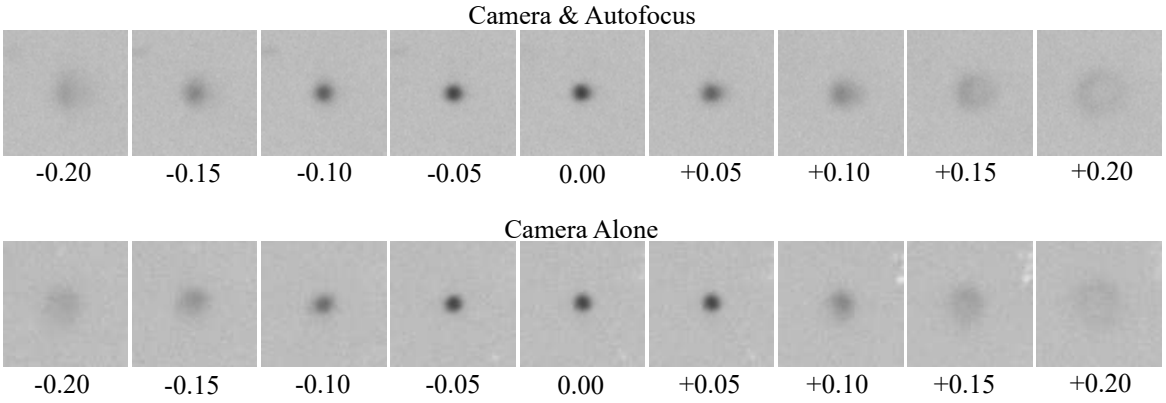


Fig. 5.4 The enlarged 10 μm dot graticule from Fig 5.3 translated in object space in 0.05mm increments, with the autofocus device upstream (top) and with the camera alone (bottom).

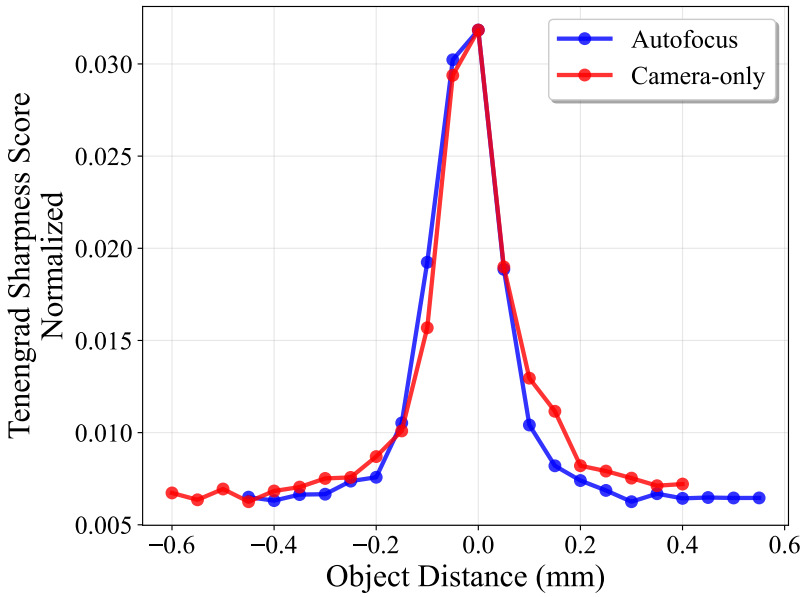


Fig. 5.5 Tenengrad sharpness scores of a 10 μm dot for the imaging camera with (blue) and without (red) the autofocus device inserted upstream. Each point represents the sharpness of the dot at varying Z-axis locations in object space, corrected for magnification and brightness differences.

We assessed whether inserting the autofocus (AF) module upstream of the imaging camera alters the system's depth of field (DOF) using a gradient-based sharpness measure (Tenengrad). A 10 μm dot graticule was translated along the optical axis in 0.05 mm steps while the two ports were operated side-by-side (camera-only control vs. camera-with-autofocus). At each axial position we computed the Tenengrad score T, defined as the mean squared Sobel gradient magnitude $\langle G_x^2 + G_y^2 \rangle$, after the usual intensity normalization. Because Tenengrad scales with spatial sampling, values from the camera-only port (which has $\approx 39\%$ higher magnification) were corrected by the square of the magnification ratio, $M^{-2} = (1/1.39)^2$, before comparison. Representative image cut-outs at matched defocus values (Fig. 5.4) illustrate that blur evolves similarly in both optical paths, with no qualitative evidence of asymmetry or aberration. Plotting the Tenengrad sharpness measure versus object distance yielded nearly identical normalised sharpness curves for the two configurations (Fig. 5.5).

Quantitatively, the AF path peaks at $z = 0.00$ mm with $T_{\max} = 3.18 \times 10^{-2}$; the camera-only path peaks at the same axial position with $T_{\max} = 1.58 \times 10^{-2}$ (after M^{-2} correction). Using a 0.9-criterion—i.e., the contiguous axial range over which $T \geq 0.9\,T_{\max}$—and linearly interpolating between the 0.05 mm samples, we obtain $\mathrm{DOF}_{0.9,\mathrm{AF}} \approx 0.069$ mm with crossings at $\Delta d = [-0.057, +0.012]$ mm about best focus, and $\mathrm{DOF}_{0.9,\mathrm{camera}} \approx 0.071$ mm with crossings at $\Delta d = [-0.055, +0.015]$ mm relative to its best focus at 0.00 mm. The peak locations therefore agree within the sampling step (≤ 50 μm), and the 0.9-widths are essentially identical (~ 70 μm).

Taken together, the normalised Tenengrad profiles are closely matched near focus: within $-0.10 \leq \Delta d \leq +0.10$ mm, the mean absolute difference between the two normalised curves is ~ 0.06 of peak ($\approx 6\%$), with the largest discrepancy occurring on the $+0.10$ mm side (about 0.20 of peak). Thus, insertion of the AF module neither compresses the DOF nor introduces directional asymmetry in the axial sharpness response.

5.4 Magnification constancy

A potential failure mode for an autofocus module that relies on a tunable air-lens is a change in magnification as the air-gap varies. To quantify any scale change introduced by the device, we imaged a bespoke precision grid graticule with 20 μm lines spaced 200 μm apart, at the two extrema of the refocus range. First, where the air-lens separation was minimal ($d_{\min} = 0$ mm, Fig. 5.6A) and maximal ($d_{\max} = 15$ mm, Fig. 5.6B). These positions correspond to extremes of the piezo-actuator's operating range, where the lens spacing differed most. Representative image fields at $d_{\min}$ and $d_{\max}$ are shown in Figure 5.6.

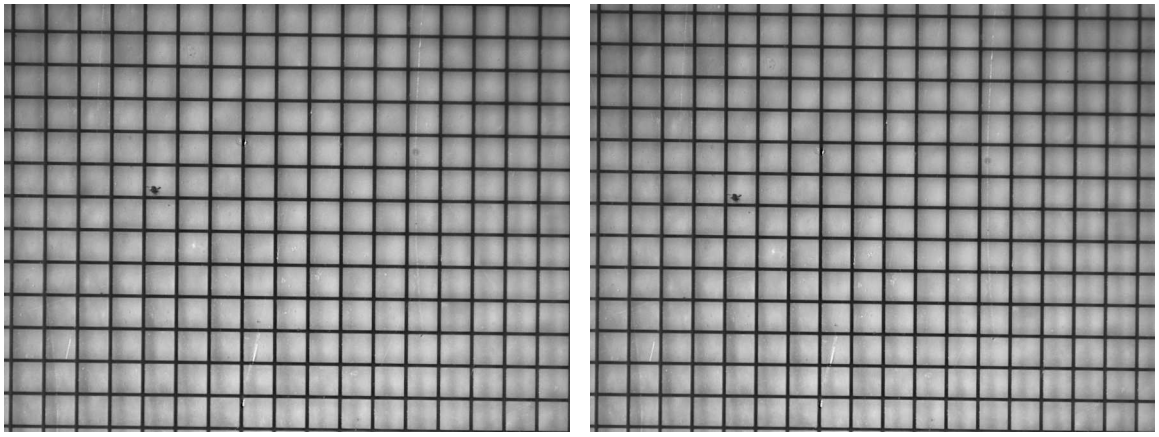


Fig. 5.6 Grid graticule with 20 μm line thickness and 200 μm spacing imaged (a) with minimum air-lens separation and (b) with maximum air-lens separation.

Counting pixels per grid pitch at each extreme yielded a relative magnification change of approximately 3.4% across the full span ($n = 3$). For context, an 8 μm erythrocyte imaged at one extreme would appear as 8.27 μm at the other ($8 \times 1.034 = 8.27$), a difference that is small relative to other sources of measurement variability.

Although small, the effect is systematic. Exact constancy could be enforced in software by rendering the frame to match the field of view to the smaller of the two extremes. For the 1944 by 1472 pixel camera used here, cropping by $\approx$32 pixels from each vertical edge and $\approx$24 pixels from each horizontal edge would be sufficient, applied as a function of focus position. This correction is not implemented in the current software. In typical angiographic recordings the convex air-lens element does not traverse the full mechanical range, so the resulting scale variation remains below the threshold of practical significance.

Having established that the device preserves magnification to within a few percent over its usable span, we next quantify the practical working range across microscopes and objectives in Section 5.5.

5.5 Working range

We next characterised the usable working range of the autofocus device. This is the span in object-space that the air-lens can accommodate across while the LoBF is held at its set-point. To measure this, the air-lens separation was swept from mechanical end-stop to end-stop (0mm to 15mm) with the controller engaged, by translating a graticule target along the Z-axis. The working range was the translation required to move the air-lens from one end to the other. Tests were performed on three microscope platforms and across eleven objectives

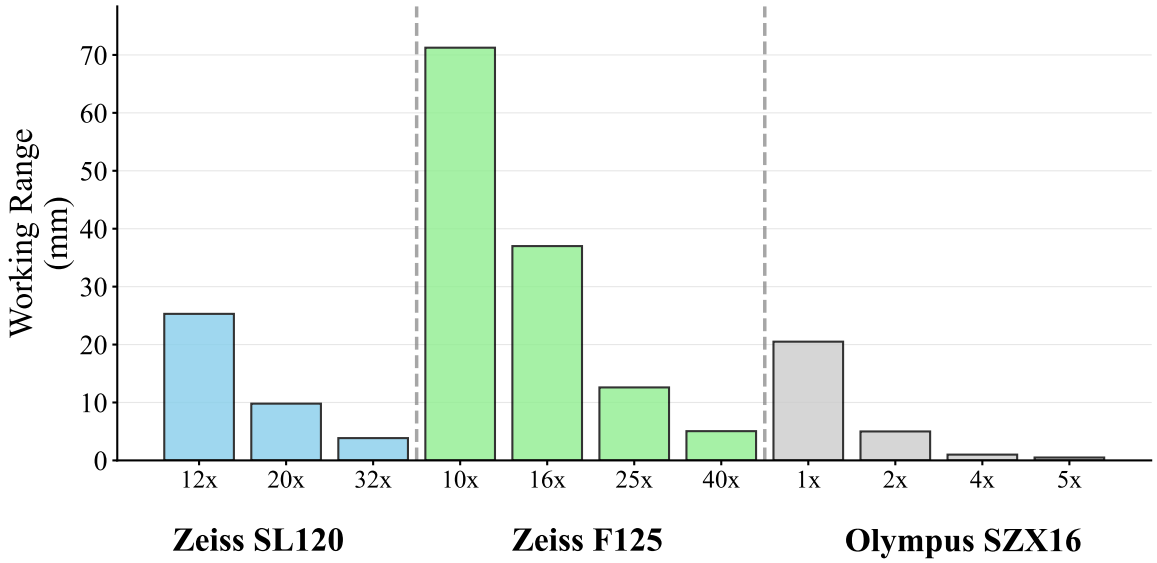


Fig. 5.7 Object-space working range for the autofocus device across eleven objectives and three microscope platforms.

in total: two slit-lamps (Zeiss SL120 in the Trumpington engineering laboratory and Zeiss F125 in the Clifford Albutt laboratory at Addenbrooke's Hospital), and a benchtop Olympus SZX16 used in the Trumpington biolab. The resulting object-space ranges are summarised in Fig. 5.7.

Within each microscope platform, the achievable range decreased monotonically with increasing magnification, as expected. However, absolute ranges varied across platforms due to differences in depth-of-fields and optical-train geometry. Overall, the working range spanned from a minimum of 0.5mm (SZX16, 5×) to a maximum of 71.25mm (F125, 10×), with a median value of 9.8mm across the eleven objectives tested. Importantly, the autofocus loop remained stably locked throughout each reported interval, demonstrating that the controller and air-lens assembly can sustain focus across the full practical travel of the device on slit-lamp and benchtop microscopes alike. Maintaining stable focus across all objectives and microscopes illustrates the inherent robustness of the control loop. This broad and repeatable working range highlights the device's adaptability and its potential as a universal autofocus add-on.

Published ranges for commercial hardware autofocus systems such as Nikon PFS, Olympus IX3-ZDC, ASI CRISP, and Zeiss Definite Focus are on the order of tens to a few hundred micrometres (e.g., PFS $\approx$ 10–100 μm; ZDC 10–200 μm; CRISP $\approx$ 10 μm; Definite Focus $\approx$ 500 μm piezo-limited). Our autofocus device has a fundamentally broader operational envelope by an order of magnitude, covering millimetres to centimetres of object space.

This highlights the difference between drift-correction designs and true long-range optical accommodation, as is achieved by our autofocus device.

5.6 Conclusion

In this chapter we moved from architectural intent to empirical validation, showing that the implemented autofocus module behaves as designed under controlled yet representative conditions. Closed-loop tests established that the controller returns the LoBF to its set-point without detectable steady-state error. Across near-step axial disturbances spanning 0.1–1.0 mm, the system settled in $t_s = 50$–142 ms at a camera-limited 60 Hz update rate, with only modest overshoot for the largest perturbations ($< 25\%$). These dynamics satisfy the speed and robustness targets derived from measured ocular motion, and indicate that focus can be held through the isolated transients and slow drifts that dominate real recordings.

Optical measurements demonstrated that inserting the module upstream of the imaging camera preserves image formation at erythrocyte scale. After correcting for the inter-port magnification and brightness differences, the point-spread function exhibited only small, isotropic broadening (AF-to-camera FWHM ratios of 1.04 horizontally and 1.08 vertically), with low and comparable eccentricity in both paths and no measurable astigmatism. Axial sharpness curves measured with a Tenengrad criterion were closely matched, with $\mathrm{DOF}_{0.9} \approx 0.07$. Thus, the module maintains the microscope's native depth-of-field characteristics and does not impose directional blur or contrast loss at best focus.

Geometric fidelity was likewise maintained. Across the full mechanical travel of the air-lens (0–15 mm), the relative change in magnification was $\approx 3.4\%$, a level that is negligible for most quantitative tasks and correctable in software if desired. More importantly, the usable working range ranged from 0.5 mm to 71.25 mm across eleven objectives on three microscope platforms, with a median of 9.8 mm. This millimetre-to-centimetre operational envelope differentiates the device from conventional drift-correction devices, which typically operate over tens to hundreds of micrometres, and underscores its suitability as a universal, long-range refocusing add-on.

Taken together, these results show that the system achieves video-rate recovery of focus, preserves optical neutrality to within a few percent, and delivers a broad, stable working range across heterogeneous optomechanical contexts. The small residual scale change and the modest overshoot at the largest step inputs are well understood and, where necessary, can be mitigated. With performance characterised and practical trade-offs quantified, we now turn from benchtop validation to end-use deployment. Chapter 6, *Application to conjunctival*

imaging and beyond, leverages these properties to hold focus despite movement artifacts across a variety of domains and microscope platforms.

Chapter 6

Application to conjunctival imaging and beyond

6.1 Introduction

Building on the system architecture and control software developed in Chapters 3 and 4, and on the performance characterization in Chapter 5, this chapter turns to *use*. Our aim is to evaluate whether a universal, video-rate, image-based autofocus module confers practical benefits under the conditions that most challenge focus maintenance.

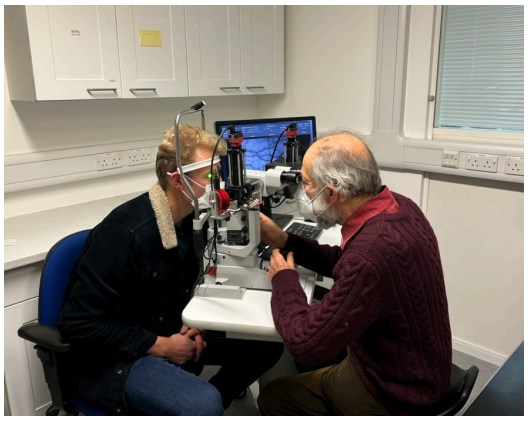
One such environment is *in-vivo* imaging of the conjunctival tissue using a clinical slit-lamp. The conjunctiva provides a stringent and clinically relevant testbed, with constant Z-axis movement, a translucent surface, and specular artefacts that erode fine microvascular detail. Section 6.2 describes the slit-lamp integration and recording protocol. We first characterise the axial motion encountered during routine conjunctival imaging by using the tilted-sensor as a calibrated rangefinder (Section 6.2.1). With this context, we then perform a paired-port comparison in which two optically identical imaging paths observe the same field simultaneously; one with autofocus enabled and one with autofocus disabled (Section 6.2.2). This experimental design isolates the effect of focus control from confounders such as illumination and temporal differences. We quantify performance using a frame-wise Tenengrad sharpness measure. Across five participants and ten trials, we compute the fraction of frames meeting a conservative in-focus criterion, and the temporal variability of sharpness. Then, Sections 6.2.3–6.2.4 outline a post-processing pipeline comprising registration, vessel segmentation, and space–time analysis along centre-lines, enabling extraction of conjunctival biomarkers such as axial flow velocity. Extracting such biomarkers is made possible by the focused videos provided by the autofocus module.

Beyond ophthalmic imaging, an image-based, video-rate autofocus module generalises to laboratory tasks where object distance varies. During in-situ microscopy of polymer sponges under tensile loading, Poisson-thinning, necking, and fixture-induced bowing change the specimen's distance from the objective, inducing blur [67]. In Section 6.3, we show that the device maintains optimal focus throughout loading and at failure, preserving microstructural detail precisely when conventional fixed-focus imaging degrades. Quantitatively, sharpness traces remain stable through large deformations while the commanded air-lens position tracks specimen deformation; qualitatively, representative frames at the moment of rupture retain clear fibre boundaries for both coarse- and fine-pore specimens.

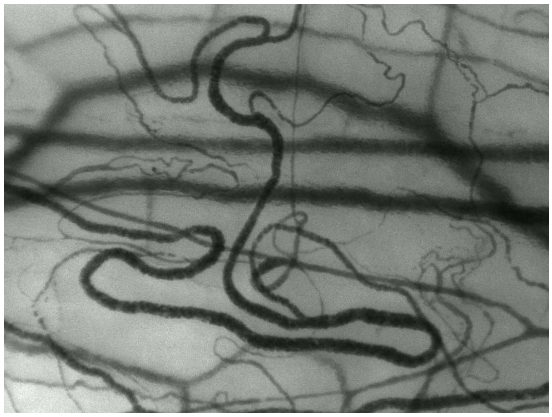
To demonstrate generality across a diverse array of imaging targets, Section 6.4 surveys performance on both bespoke geometric graticules and representative real-world specimens undergoing dynamic axial motion. We assembled a diverse array of imaging targets: fine and coarse grids, line gratings, dot fields, a multilayered microvasculature phantom, a printed-circuit board, and a human hair. In almost all cases, autofocus held a user-selected region in focus while fixed-focus controls blurred as the target was perturbed.

Across these three experiments, we assess technical performance and the extent to which the design goals of speed, accuracy/robustness, and clinical operability are achieved. Ultimately, we show that the proposed architecture delivers practical, quantitative benefits across a diverse array of clinically realistic settings.

6.2 Conjunctival imaging on a slit-lamp



(a) Autofocus device mounted on a Zeiss SL120 slit-lamp set-up for HVI.



(b) A representative frame taken using HVI.

Fig. 6.1 Experimental setup for HVI and a sample frame taken from Supplementary Video S2.

The conjunctiva is an attractive site for non-invasive, repeatable assessment of the microcirculation with relevance to both ocular and systemic disease [52, 72, 80]. To maximise erythrocyte contrast, we employ Hemoglobin Video Imaging (HVI) on a Zeiss SL120 slit-lamp mounted on a stabilised desk. HVI inserts a 505–575 nm band-pass filter into the illumination path and a hot mirror that rejects > 730 nm. This places the illumination source near a haemoglobin absorption peak so that erythrocytes appear as dark dots against a bright scleral background [79]. In practice, the patient's chin and forehead are braced against a stabilised desk while the white of the eye is illuminated and imaged, allowing single-erythrocyte resolution at video rate without exogenous dyes or mydriatic drops. A representative field is shown in Fig. 6.1b. In routine *in-vivo* slit-lamp imaging, however, this resolution is constantly tempered by optical artefacts and movement artefacts from microsaccades, heart action, and respiration. These effects erode erythrocyte sharpness, complicate registration, and make space–time analyses impossible unless focus is actively maintained.

To view the image, the operator may either use the eyepiece or the attached Alvium 1800 U-291m imaging camera, which is streaming to a computer screen via the autofocus software. The camera and eyepiece share an identical focal plane; the physical arrangement is shown in Fig. 6.1. Because axial motion is the dominant failure mode during conjunctival recording, the next subsection quantifies the typical Z-excursions encountered under routine conditions (Section 6.2.1).

6.2.1 Typical motion

Quantifying axial patient motion was necessary to understand the control requirements of the autofocus mechanism during routine conjunctival imaging on the slit-lamp. To do so, we repurposed the tilted-sensor autofocus module as a passive rangefinder by disabling the air-lens (no active refocusing), allowing the imaging camera to go out-of-focus while the algorithm reported the per-frame LoBF along the tilted sensor. A calibration step following the experiment mapped this sensor coordinate to physical displacement: a graticule mounted on a micrometre was stepped along the optical axis in known increments, and the same LoBF readout was recorded. The mapping was approximately linear, yielding slopes of 110.7 pixels/mm at $12\times$ magnification ($n = 5$) and 231.2 pixels/mm at $20\times$ ($n = 5$). With this calibration, the best-focus coordinate during *in-vivo* recordings provided a frame-wise estimate of the patient's Z-axis motion relative to the plane of maximal sharpness.

We repeated the protocol three times as the system matured and report the latest data here. In June, eight eyes from four participants were recorded at both $12\times$ and $20\times$ magnification. A representative $20\times$ trace is shown in Fig. 6.2, which exhibits sub-millimetre to

millimetre-scale excursions over an $\sim$80 s interval. Features consistently observed across traces included:

- **Blink artefacts**: large, brief spikes in the axial trace, corresponding to eyelid closure and recovery.
- **Drift to tilted-sensor bounds**: slow excursion of the LoBF to the edge of the tilted sensor ("beyond bounds"), which saturates the rangefinder but would be unproblematic when active refocusing is enabled.
- **Transient jumps**: abrupt axial displacements that dominate defocus in HVI videos; the worst event shown here was on the order of 0.7 mm over 1.3 s, but the worst observed was around 0.8 mm over 1.2 s.

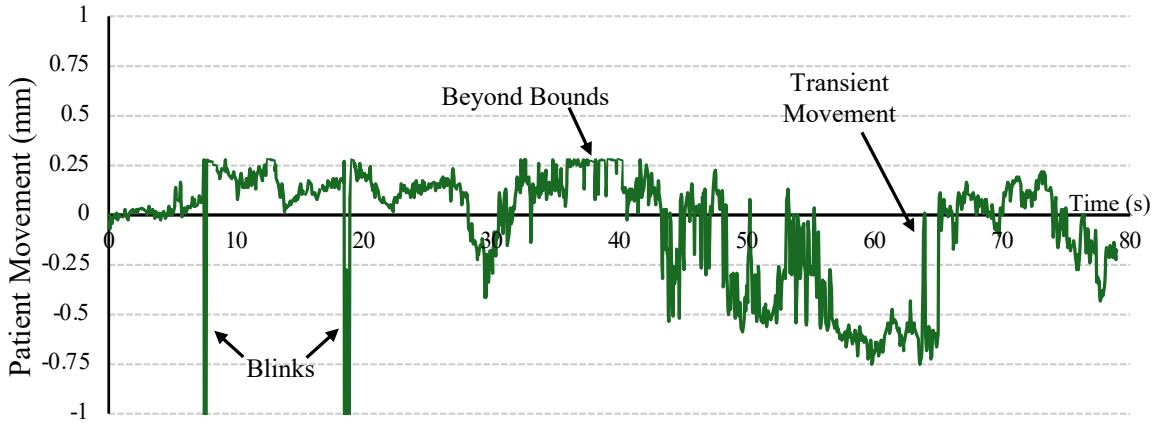


Fig. 6.2 Typical patient movement during a conjunctival angiogram.

To probe whether movement artifacts at a persistent frequency were detectable, we analysed patient movement with a discrete Fourier transform. Across trials, no reproducible peaks emerged from the noise floor; energy was concentrated below 1 Hz and in aperiodic components. This is consistent with motion being governed by low-frequency drift and intermittent transients, with any respiratory ($\sim$ 0.3 Hz) or cardiac ($\sim$ 0.7 Hz) modulation present only weakly. Together, these observations suggest that autofocus must reject slow drift yet respond quickly to step-like axial perturbations. We therefore next compare performance with and without autofocus using a paired-port experimental design (Section 6.2.2).

[**?**]

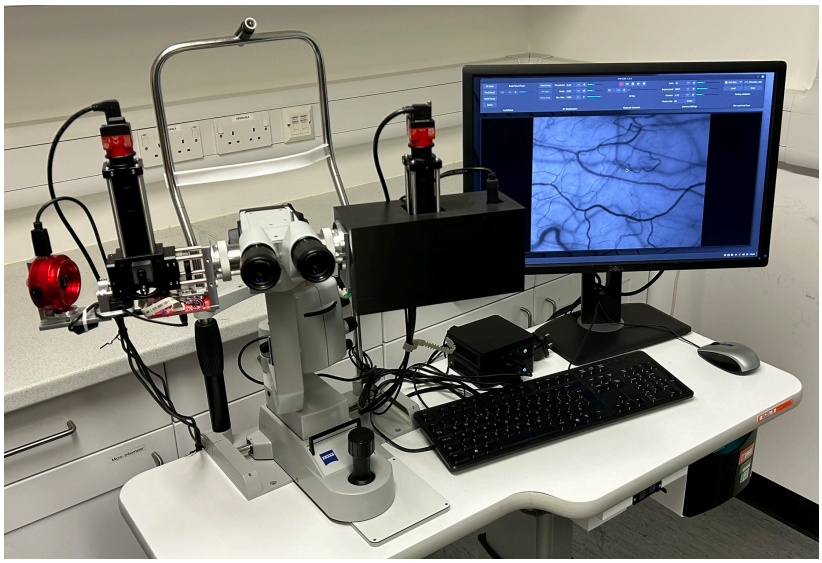


Fig. 6.3 Experimental paired-port setup using a Zeiss SL120 slit-lamp with prototypes on either camera port; one has autofocus enabled, the other has autofocus disabled.

6.2.2 Performance with and without autofocus

To isolate the effect of focus control from confounders, we implemented a paired-port experiment on a clinical slit-lamp (Zeiss SL120) where two optically identical autofocus prototypes were mounted simultaneously, each feeding an identical Alvium 1800 U-291m camera (Fig. 6.3). Both imaging paths were shrouded with matched black covers to suppress ambient light; one prototype operated with the air-lens enabled (autofocus), the other with the air-lens disabled (fixed focus). The slit-lamp magnification was set to $32\times$, and both cameras were run at gain 15, exposure 8000 μs, and gamma 1.1. Videos were acquired on either the left or right temporal ocular surface from six participants, for a total of twelve collected trials. Capture occurred near-simultaneously and time alignment was done by matching blinks and distinctive features between the videos, to ensure similar illumination and subject fatigue conditions across ports. One participant's data was excluded due to unusually excessive blinking. Each trial nominally lasted between 20 s to 30 s. After removing blink-corrupted frames, the representative trial shown below yielded ≈ 17.5 s of analysable data.

We quantified performance using a frame-wise Tenengrad sharpness metric and two summary statistics: (i) the fraction of frames classified as in-focus, using a conservative criterion $T \geq 0.9\,T_{\max}$ within each trial, and (ii) the temporal standard deviation of Tenengrad sharpness as a measure of focus stability. The $T \geq 0.9\,T_{\max}$ criterion was informed by depth-of-field testing in Section 5.3.2. Across trials, autofocus increased the time spent in focus and reduced sharpness variability. The fraction of frames meeting the in-focus criterion was

$88.0 \pm 8.8\%$ (mean $\pm$ s.d.) with autofocus compared with $52.9 \pm 30.3\%$ under fixed focus ($n = 10$) (Fig. 6.7 b). A paired t-test revealed a significant 35% mean increase in focused-fraction with autofocus ($t_9 = 3.06$, $p = 0.014$), indicating that the autofocus trials significantly outperform the fixed focal plane trials. In parallel, the temporal standard deviation of Tenengrad sharpness decreased from $(5.95 \pm 4.81) \times 10^6$ (fixed focus) to $(4.22 \pm 0.82) \times 10^6$ (autofocus), indicating more stable focus over time (Fig. 6.7 a). These results demonstrate that the autofocus system fulfils its design objectives of speed and accuracy, and successfully improves focus during dynamic imaging.

Qualitatively, the autofocus device maintained stable clarity of conjunctival microvasculature despite axial drift and blink-related transients. Without autofocus, distinct erythrocytes and fine capillary loops intermittently softened or disappeared across all trials. With autofocus, capillary loops, vessel bifurcations, and blood flow remained crisp, revealing erythrocytes and occasional leukocytes translating along vessel centre-lines and margination zones (Fig. 6.4). Supplementary Video S3 provides a side-by-side comparison of autofocus and fixed-focus imaging of the same microvasculature. Both fields of view are XY-registered so that only axial (focus) differences remain. Supplementary Videos S1 and S2 show 40 s and 60 s continuous in-focus recordings of the conjunctiva using the microscope device, stabilised in XYZ axes. Focused videos of this duration were previously unattainable. Single-erythrocyte resolution was consistently observed at both $20\times$ and $32\times$ magnification on the slit-lamp.

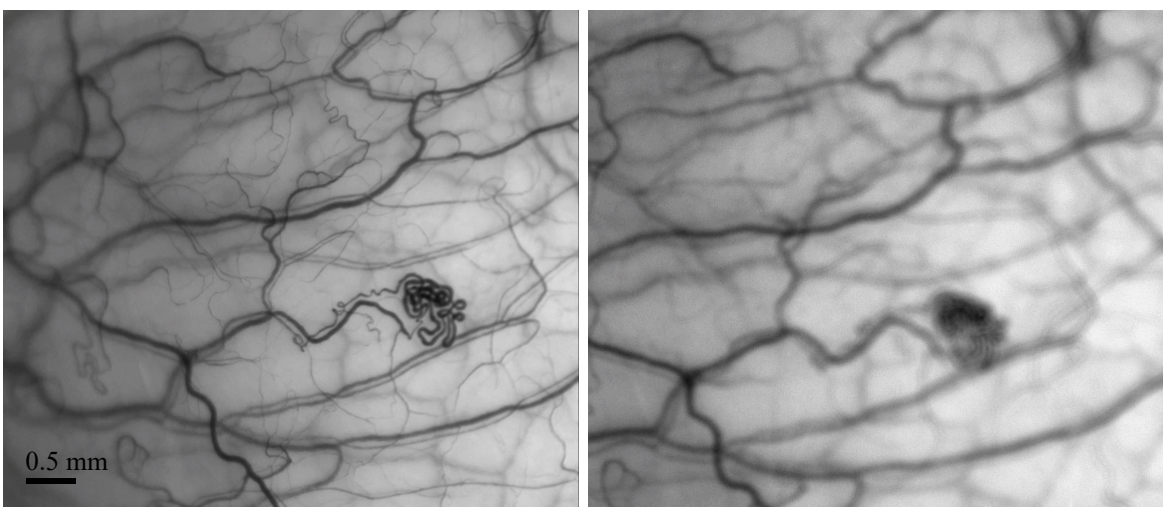
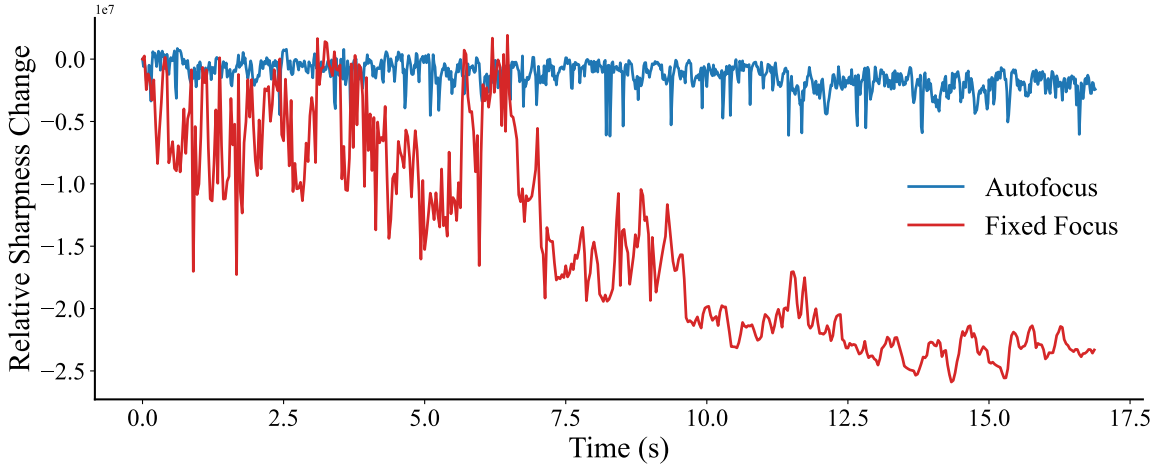
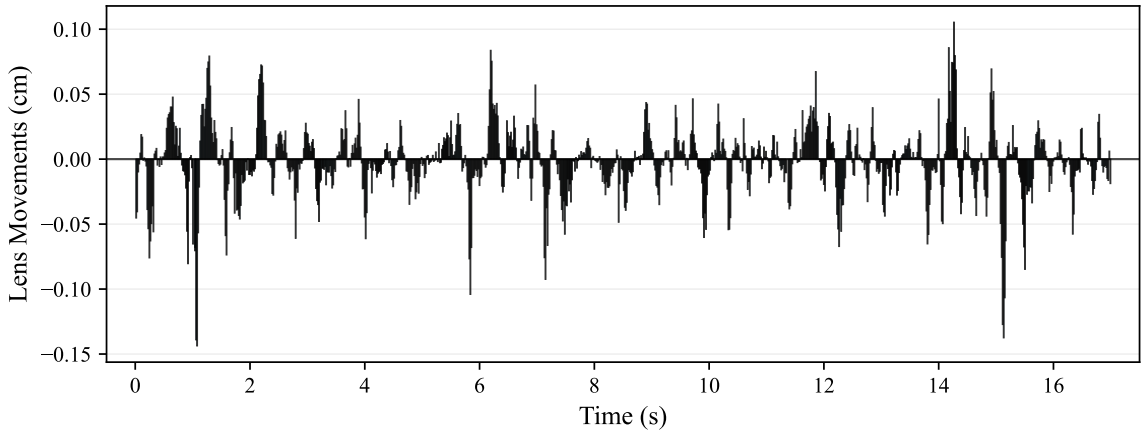


Fig. 6.4 A sample frame after nine seconds of an angiogram taken simultaneously with and without autofocus active. The full video with both autofocus and fixed-plane perspectives is shown in Supplementary Video S3.



(a) Relative sharpness change over time for autofocus (blue) and fixed-focus (red) conditions. Trial lasted 20s; blink data is removed.



(b) Axial lens displacements commanded by the autofocus controller during the same trial.

Fig. 6.5 Sharpness and lens instruction data during a representative paired-port trial.

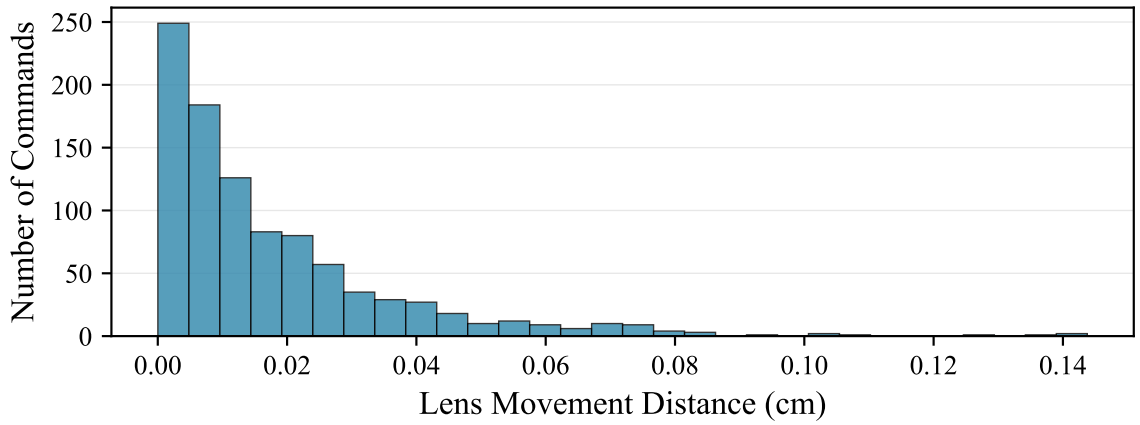
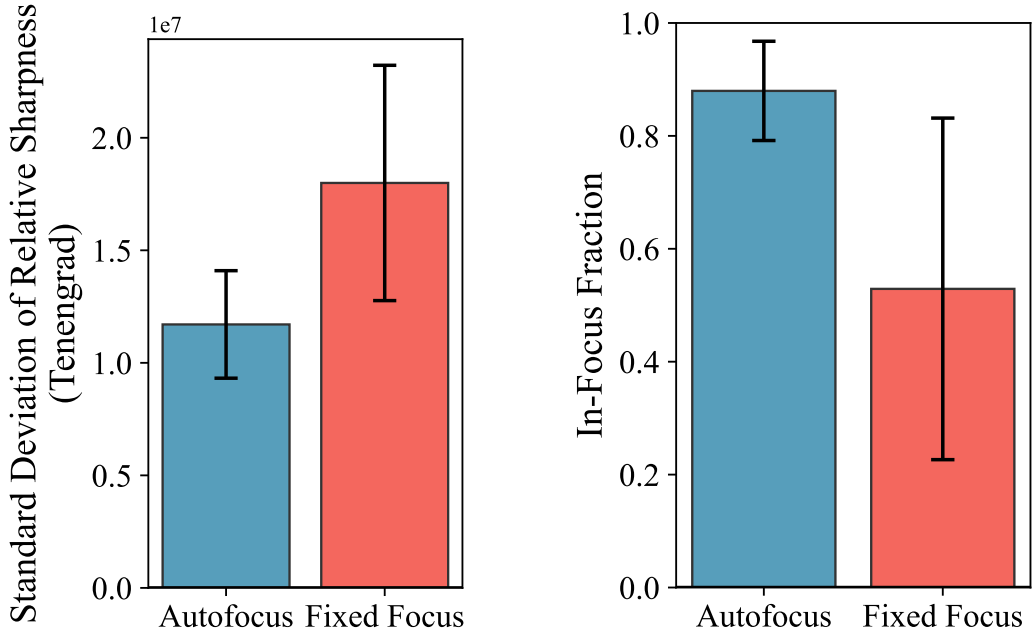


Fig. 6.6 Distribution of commanded lens movement magnitudes across the paired-port experiment.



(a) Standard deviation of the Tenengrad sharpness metric across the trial. Lower variation indicates more stable focus.

(b) Fraction of frames meeting the focus criterion ($T \geq 0.9T_{\max}$). Higher values indicate more time spent in focus.

Fig. 6.7 Summary statistics of the paired-port experiment comparing autofocus and fixed-focus operation across five subjects and ten trials.

The impact of the autofocus system is illustrated by the representative, blink-trimmed trial shown in Fig. 6.5 a. The autofocus sharpness trace remains sharp and comparatively flat relative to the fixed-focus trace. Concomitant axial air-lens commands (Fig. 6.5 b) show frequent but small displacements as the controller compensates for ocular Z-motion. Over the entire experiment, 959 non-zero movement commands were issued; 93.2% (894/959) were within ± 0.05 cm, consistent with a controller that primarily performs fine corrections rather than large excursions (Fig. 6.6). The speed and precision of these lens adjustments indicate a fast and accurate sensing-to-accommodation loop.

Taken together, these results show that, for clinically realistic conjunctival imaging, the autofocus module materially improves image quality and its temporal stability relative to a fixed focal plane. The improvement is evident quantitatively, shown by higher in-focus fraction and reduced sharpness fluctuations. It is also evident qualitatively, where microvascular features and individual erythrocytes remain distinguishable despite patient motion. The paired-port, time-synchronised experimental design isolates the contribution of focus accommodation. The narrow distribution of air-lens movements indicates precise, small-amplitude corrections that track physiological axial motion closely. In practical terms, maintaining focus in this way enables downstream processing (registration, vessel segmentation, and centre-line analysis), which is discussed in the following Section 6.2.3.

6.2.3 Video post-processing

The previous Section 6.2.2 delivered long (20–60 s) in-focus recordings of the conjunctival microcirculation with single-erythrocyte resolution. Two such videos are shown in Supplementary Video S1 and S2, following post-processing steps discussed here.

Our acquisition software records raw image sequences as full-resolution grayscale `.pgm` files and offers real-time image registration on playback. However, offline registration proved more accurate for publication-quality output and for post-hoc video analysis. To prepare the supplemental videos, we first removed blinks via a semi-manual pass, then applied minor contrast and sharpening adjustments so that the available dynamic grayscale range was fully utilised. Next, we performed affine image registration in ImageJ on the blink-free sequences. Blinks otherwise can cause registration failure, especially at higher magnifications. Finally, we compressed the frames to MP4 using FFMPEG commands optimised for minimal loss. This pipeline yielded the videos supplied with the paper, each stabilised in XYZ, demonstrating preserved single-erythrocyte resolution. While this workflow produces visually faithful videos suitable for viewing and archiving, extracting quantitative biomarkers requires a more in-depth analysis pipeline. This is presented in the next section.

6.2.4 Pipeline for biomarker extraction

Access to repeatable videos of the conjunctival microcirculation with single-erythrocyte resolution prompts analysis efforts through biomarker extraction. Therefore, this subsection summarises a pipeline that extracts microvascular biomarkers from haemoglobin video imaging (HVI) sequences. This work was developed in tandem with two fourth-year students, and is included here to indicate the direction of travel for this thesis rather than as its central contribution. In brief, autofocus-stabilised frames from the slit-lamp are (a) registered and pre-processed, (b) combined into a minimum-intensity image (MII), (c) segmented using a U-Net algorithm, (d) skeletonised to define centre-lines and local diameters, and (e-f) analysed with space–time image velocimetry (STIV) using kymographs to estimate velocity profiles, from which network-level velocity maps are assembled (Fig. 6.8).

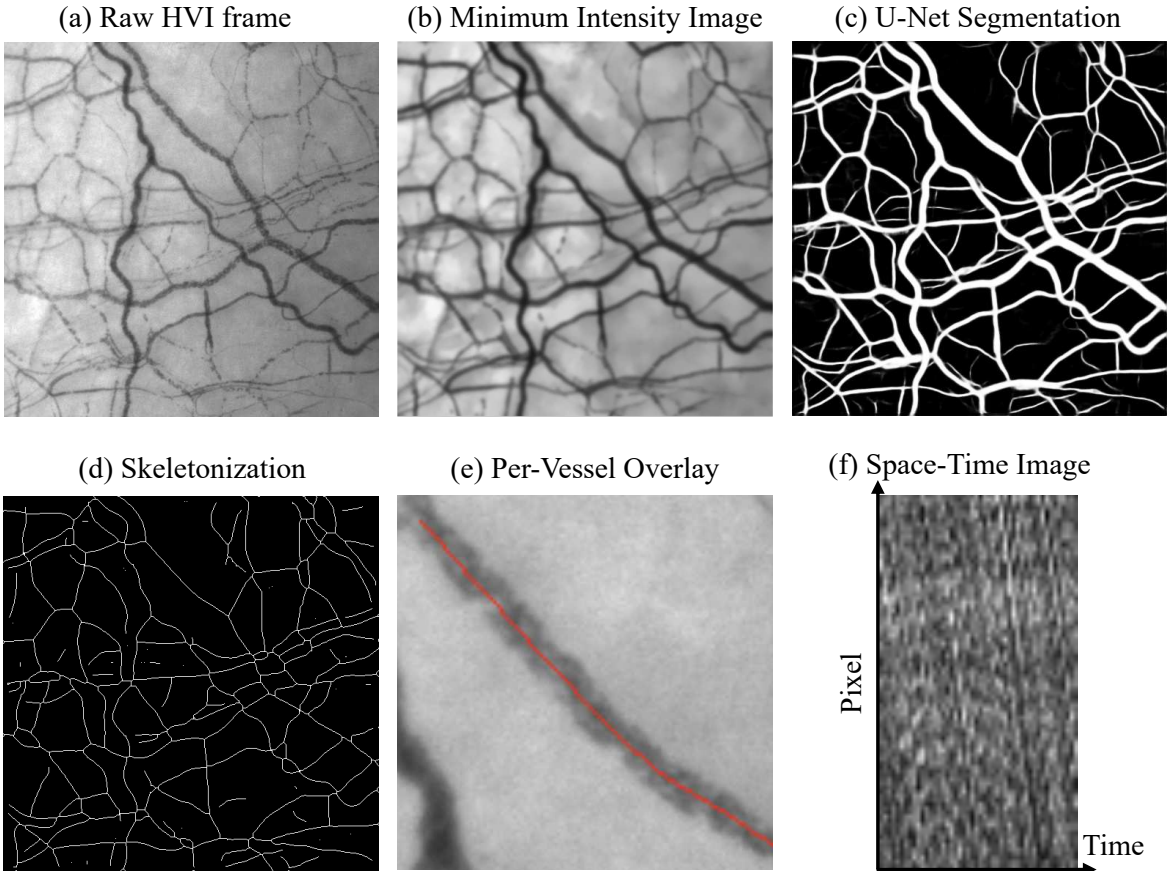


Fig. 6.8 Processing pipeline from raw HVI sequence to extraction of various biomarkers. (a) Raw image; (b) minimum-intensity image (MII); (c) attention U-Net segmentation (d) skeletonization; (e) per-vessel overlay; (f) space-time image (STI). Panels adapted from an undergraduate student project [54].

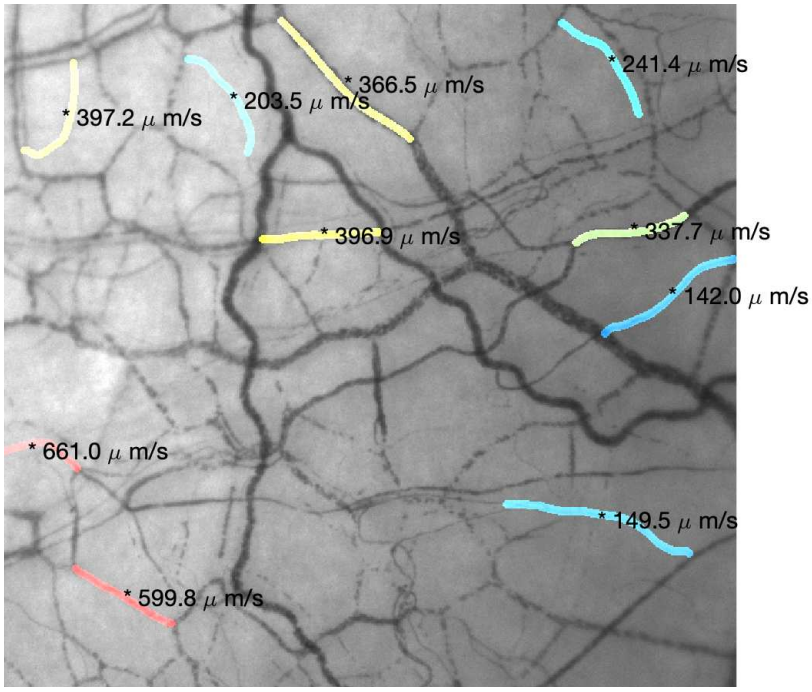


Fig. 6.9 Heat-map of velocities across various vessels. Panel adapted from an undergraduate student project [54].

Step-by-step, sequences of raw frames (Fig. 6.8a) are stabilised by multi-pass rigid-body alignment (translation+rotation) using a StackReg/PyStackReg implementation, with optional cropping between passes to reduce boundary artefacts. Affine transforms and ECC-based refinements were trialled, but rejected for routine use because small shearing tended to introduce vessel aspect-ratio distortions. The simpler rigid model offered a better bias–variance trade-off for flow estimation. Frames are then denoised and contrast-enhanced. Among filters explored, Gaussian blurring provided light high-frequency suppression without erasing intravascular texture required for successful STIV. By comparison, non-local means reduced background noise more aggressively, but over-smoothed valuable erythrocyte clump patterns. Therefore, contrast-limited adaptive histogram equalisation (CLAHE) was adopted (tile size $\sim 20 \times 20$ px) to improve vessel–tissue separation.

Following registration and pre-processing, a minimum-intensity image (MII; Fig. 6.8b) is computed by aggregating the stabilised stack. This step eliminates discontinuities in the capillary outline seen in a single HVI frame. The MII then serves as a robust starting point for segmentation. Vessel segmentation (Fig. 6.8c) is performed with a convolutional neural network variant called U-Net, equipped with additional spatial attention ("SA-U-Net"). A fuller-capacity FR-U-Net pre-trained on retinal datasets was also evaluated; although acceptable on superficial vessels, it struggled with the layered conjunctival circulation. By

contrast, the attention U-Net offered improved discrimination of deep versus superficial beds and CPU-feasible inference. To reduce annotation burden, a semi-automated protocol was used to fine-tune only the final layers on a small labelled HVI set, yielding a binary mask suitable for downstream geometry and flow extraction.

With the binary mask from the segmentation step, we skeletonise and prune at branch points to isolate centre-lines (Fig. 6.8d). Local vessel diameters are then estimated by a Euclidean distance transform from centre-line to boundary. This yields diameter as a function of arclength for each segment.

For each centre-line, we assemble a kymograph (or space–time image, Fig. 6.8f) by sampling along the path across consecutive frames. In the original implementation, the dominant orientation of texture in the STI was estimated from the log-magnitude of its 2-D spatial FFT by fitting a line through the principal frequency components. In the later student's update, orientation is instead obtained by maximising the mean-squared response of the STI to a bank of oriented Gabor filters; a coarse-to-fine search over angles provides the dominant slope and, by spatial/temporal calibration, the axial velocity. The Gabor-energy approach proved more tolerant to noise and processing artefacts than the FFT fit and, importantly, supports subdivision of the STI into patches to recover secondary orientations (e.g. local velocity changes or brief flow reversals). Aggregating per-segment estimates produces a per-vessel overlay (Fig. 6.8e) and a velocity heat map over the segmented network.

To move beyond centre-line velocities (not shown in Fig. 6.8), parametric B-splines are fit to centre-lines and translated along local normals to generate a set of equidistant tracks spanning the diameter of a vessel. Kymograph analysis on these offsets yields a transverse velocity profile. Combining profiles with diameters gives relative flow rates $Q \propto \bar{v} D^2$ along a segment. In several exemplary vessels, spatial changes in texture orientation corroborated junctional acceleration and aided discrimination between true junctions and projected overlaps in the layered circulation. Deviations from simple flow-continuity scalings were also observed, consistent with known microvascular complexities (non-Newtonian rheology, compliance, noncircular cross-sections).

The biomarkers yielded by this pipeline include: (i) per-segment axial velocity V_a along the vessel centre-line; (ii) cross-sectional velocity V_c estimated from space–time images sampled across the vessel diameter, including off-centreline regions toward the vessel edges; (iii) local vessel diameter $D(s)$ at any point along the centre-line arclength s; (iv) a network-level velocity heat map combining all per-segment V_a estimates; (v) morphological descriptors derivable from the segmentation mask (average segment length, branch density); and (vi) optional local velocity-change metrics $\Delta V_a/\Delta t$ from patch-wise kymograph analysis (e.g. secondary flow orientations or reversals). From these quantities, derived haemodynamic

parameters such as flow rate $Q = V_c \pi D^2/4$, wall shear rate WSR $= 8V_c/D$, and wall shear stress WSS $= \mu$ WSR can subsequently be computed, following the relationships outlined in Section 2.4.2.

At present, analysis is offline and semi-manual, although aided by a bespoke GUI. Accuracy is constrained by limited HVI-specific annotations and by residual motion in challenging clips. Future work will further automate the processing pipeline, expand labelled datasets for improved learning-based processing, and formalise quality-control metrics (e.g. reject segments with poorly-resolved erythrocytes). Most importantly, prospective clinical validation is required to assess the repeatability and diagnostic value of each biomarker. These early results indicate that autofocus-stabilised video enables reliable extraction of quantitative conjunctival biomarkers, establishing a foundation for future clinical validation and diagnostic studies. [**?**]

6.3 Polymer sponge tensile test on a benchtop microscope

Fig. 6.10 The autofocus device mounted on the Olympus SZX16 stereomicroscope.

Beyond slit-lamp microscopy and conjunctival imaging, we sought to test whether the autofocus system can maintain focus on different microscope platforms. In-situ bench-top microscopy of polymer sponges under tensile load offer a convenient yet stringent experiment

where objective distance changes continuously [67]. Poisson-thinning, localised necking, and fixture-induced bowing alter the axial position of the specimen during deformation, producing blur in fixed-focus recordings [67], especially at tensile failure. We mounted the autofocus system on an Olympus SZX16 stereomicroscope using the microscope interface from Section 3.7, and shown in Figure 6.10. We operated the microscope at $2.5\times$ magnification and recorded with camera settings of exposure 8000 μs, gamma $= 1.1$, and gain $= 20$. Notably, although the architecture in earlier sections assumes a collimated camera beam, this microscope's port is a *converging* beam. While collimation can be introduced with a simple lens (as is done in Section 5.5), the present experiment used the default converging relay to probe robustness. We observed no appreciable magnification change introduced by the autofocus actuation below 0.2cm, and accommodation was successful. Because the SZX16 provides only a single camera port, autofocus and fixed-focus trials were acquired sequentially rather than simultaneously.

Two different sponges were imaged from strain onset to tensile failure. Both the coarse- and fine-pore sponge are shown in Figure 6.11 at tensile failure, for both the autofocus and fixed-focus trials. For the coarse-pore sponge, Supplementary Video S4 presents the autofocus and fixed-focus recordings side-by-side, with XY registration applied to both recordings. Qualitatively, the difference at the point of failure is stark for both pore scales. The representative frames show that the fixed-focus trials are visibly defocused at tensile failure. By comparison, the autofocus trials clearly preserve distinct fibre boundaries and microtexture at the same instant (Fig. 6.11). These examples illustrate the practical benefit of maintaining focus precisely when rapid deformation would otherwise cause defocus on the features of interest.

Quantitatively, the Tenengrad sharpness traces corroborate the visual observations. Prior to loading, both autofocus and fixed-focus trials exhibit comparable sharpness baselines; upon strain onset (vertical marker), the fixed-focus trace loses sharpness and remains so to the end of the trial. By comparison, the autofocus trial sustains high sharpness throughout the deformation and tensile failure (Fig. 6.12). This behaviour is consistent across repeated tensile pulls.

The commanded position of the tunable air-lens tracks the specimen's axial motion smoothly, as shown by Fig. 6.13. After strain onset at approximately 5 s, the lens translates monotonically, reaching ~ 0.1 cm (about 1 mm) by the onset of failure (around 12 s), and then relaxes as the specimen unloads (Fig. 6.13). Together with the sharpness results, this indicates that the controller accommodates millimetre-scale object-space motion without focus hunting or instability, even in a non-collimated microscope port.

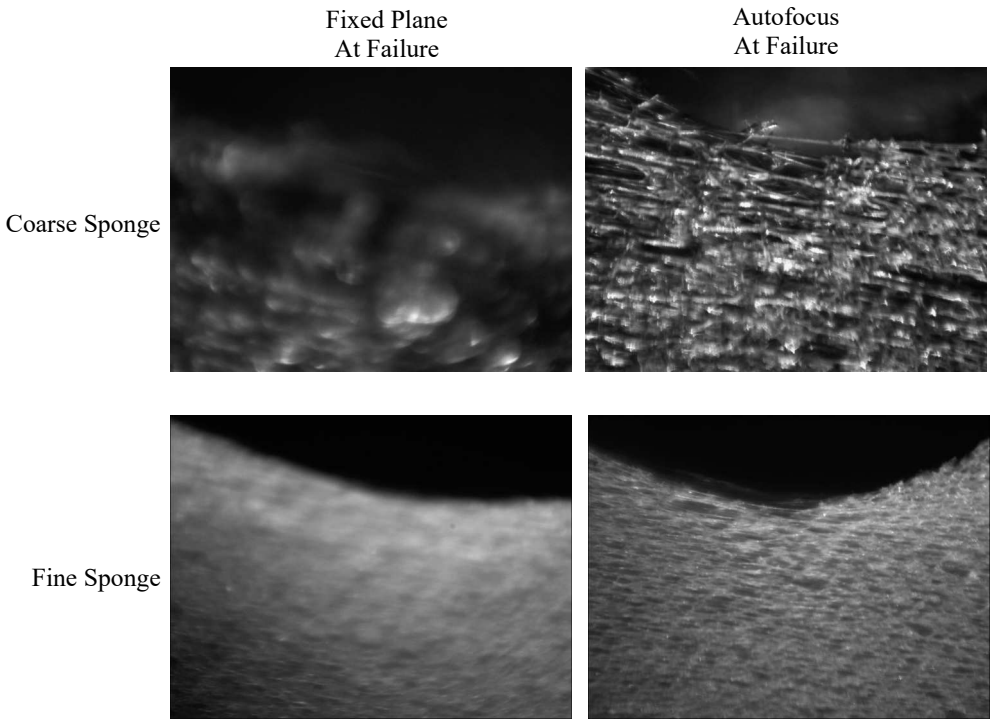


Fig. 6.11 Representative frames at the point of mechanical failure for two sponge specimens imaged under fixed-focus and autofocus conditions.

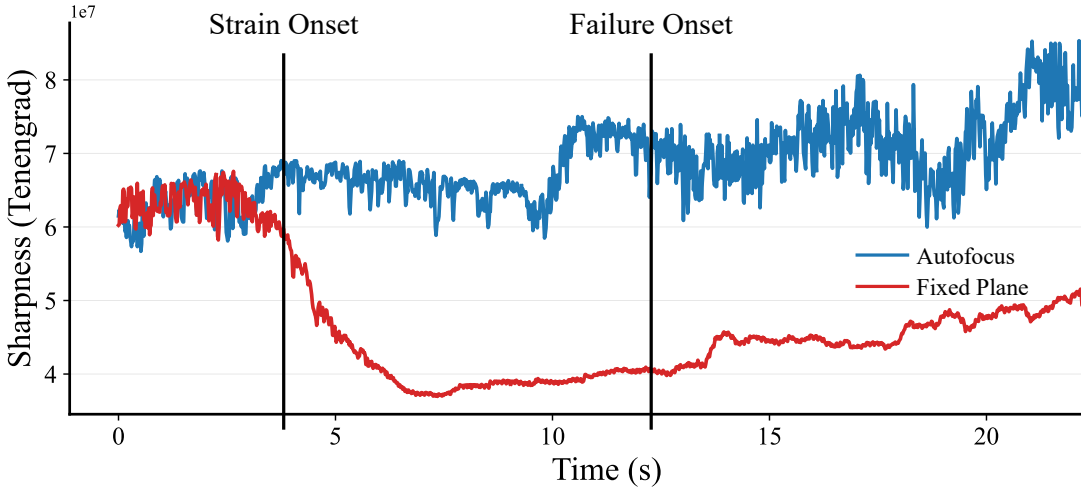


Fig. 6.12 Video sharpness during the *in-situ* coarse sponge tensile test with autofocus (blue) and fixed-plane (red) trials. Vertical markers approximate the onset of strain and of structural failure.

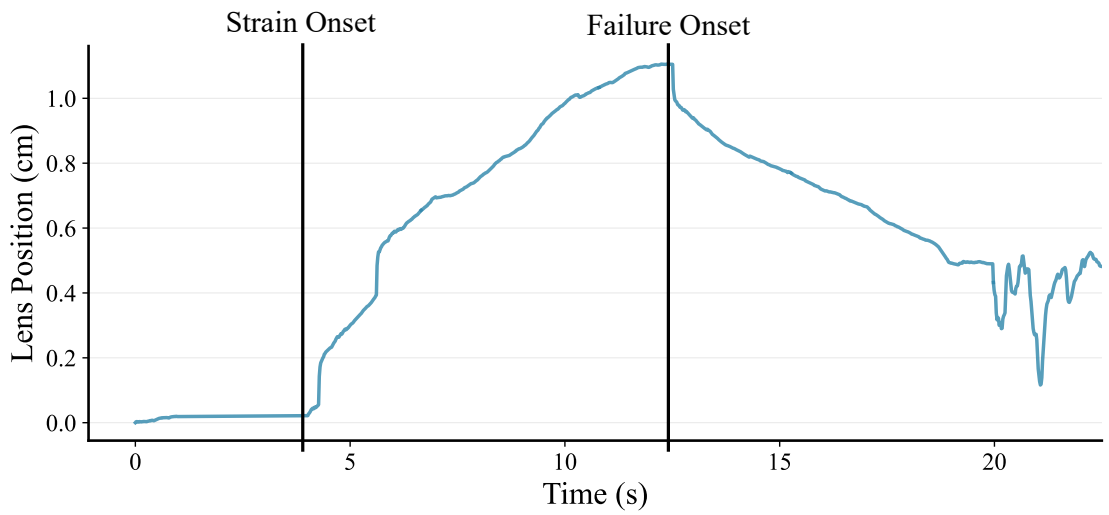


Fig. 6.13 Axial position of the tunable air-lens holding focus during the autofocus trial from Fig. 6.12.

In sum, the polymer-sponge test demonstrates that the autofocus system preserves image clarity under substantial, dynamic perturbations in benchtop microscopy. This is evidence that its utility extends well beyond the slit-lamp, and beyond the application of conjunctival imaging. Having verified platform independence, we next evaluate the system's performance across a diverse range of bespoke and realistic imaging targets (Section 6.4).

6.4 Other imaging targets

To stress–test generality beyond conjunctival recordings, we assembled a deliberately heterogeneous panel of targets comprising fine and coarse grids, line gratings, dot fields, a multilayered microvasculature phantom, a printed-circuit board (PCB), sponges (from the previous section), and a human hair. Grids, lines, and dot fields were printed with line width sizes ranging from $10-500$ μm. In total, we printed 65 bespoke graticules and prepared four object-space realistic targets. Targets were mounted on the headrest of the SL120 slit-lamp and subjected to a manual sinusoidal perturbation predominantly along the optical (Z) axis. For each target we acquired matched videos with autofocus enabled and disabled, holding identical acquisition parameters (exposure 8000 μs, (gamma 1.1), gain 15). A user-selected region of interest (ROI) defined the autofocus lock point, which typically maximised sharpness across the image, although not always (as shown in Supplementary Video S6). Representative frames at or near the worst defocus excursion are shown for a selection of imaging targets in Figure 6.14.

Qualitatively, across the bespoke geometric targets (Figure 6.14a), autofocus retained focus and high-frequency structure, whereas fixed-plane imaging visibly blurred. Supplementary Video S5 demonstrates a representative side-by-side video, using the PCB's dynamic test filmed with and without autofocus. Both perspectives are post-processed with XY registration, so that the only difference is the Z-axis autofocus when the air-lens is enabled. With autofocus enabled, the PCB remains in-focus, whereas it loses focus regularly without autofocus engaged.

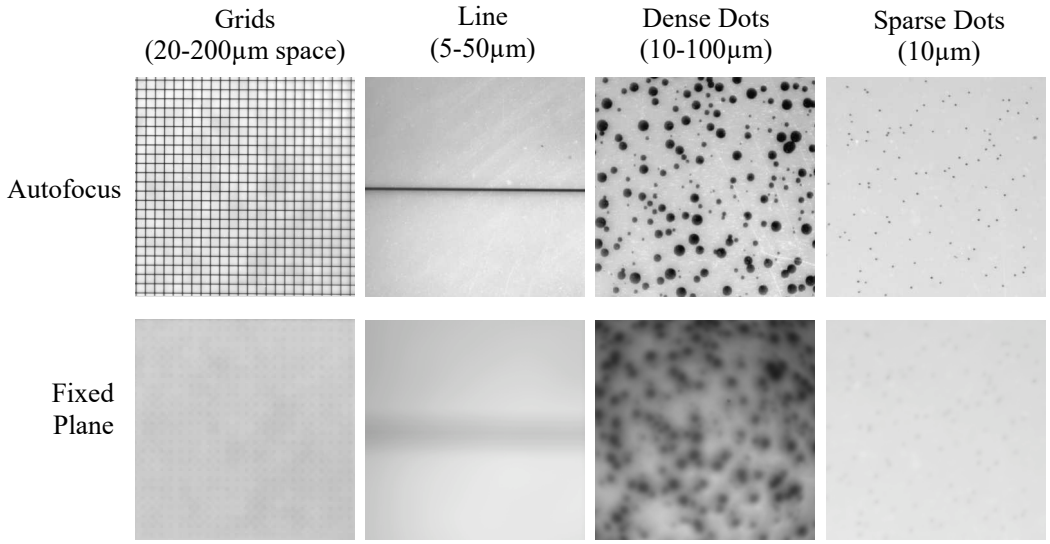
These qualitative impressions are corroborated by the time-resolved Tenengrad sharpness traces, with one representative trial shown in (Figure 6.15). Under identical perturbation, the autofocus condition sustains a near-constant sharpness level, whereas the fixed-focus control exhibits periodic collapses in sharpness as the sample traverses the focal plane. Such a sharpness trace was observed across almost all trials.

A notable limitation emerges on sparsely textured anisotropic targets. In particular, the line graticule, when oriented vertically or near-vertically, confounded the autofocus system because there was effectively no horizontal texture from which to build a valid sharpness curve. By contrast, the line graticule oriented horizontally or diagonally enabled autofocus as normal. The human-hair preparation did not suffer this failure mode, because the feature occupied a sufficient fraction of the image. In practice, when imaging an object that lacks texture across a sufficient horizontal portion of the field, adding a grid graticule as a background or cover slip resolves the issue. Future prototypes could also further mitigate such cases by adopting more complex orientations of the tilted sensor to furnish robust focus cues irrespective of scene anisotropy (discussed further in Chapter 7).

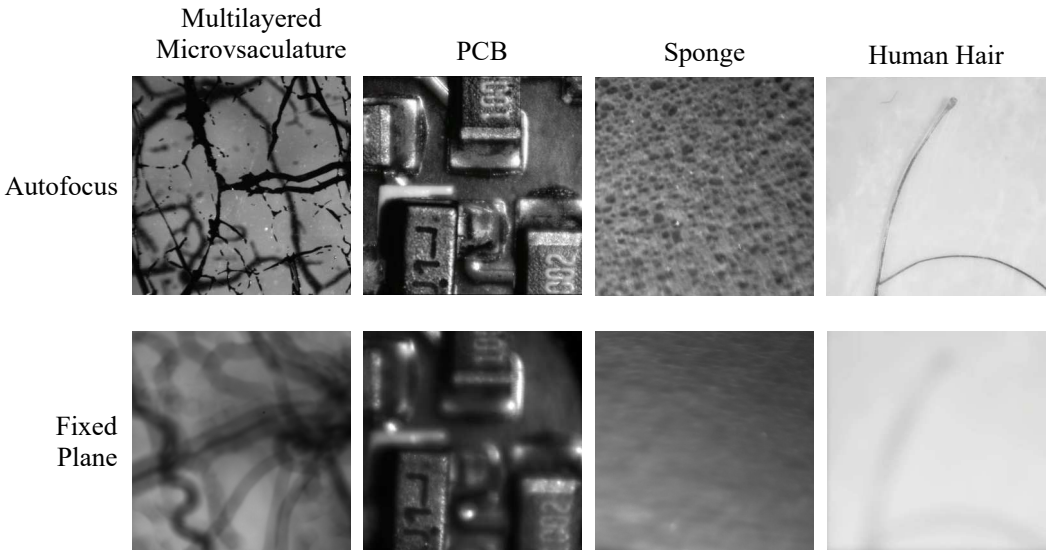
This experiment establishes that the autofocus module operates effectively across both engineered and natural targets, sustaining image clarity in any ROI despite substantial differences in texture, structure, and target depths. These findings underscore the system's universality and its suitability for deployment beyond ophthalmic imaging.

6.5 Conclusions

This chapter demonstrates that the autofocus module delivers practical value in its primary clinical setting: *in vivo* conjunctival imaging on a slit-lamp. By quantifying typical axial excursions (Section 6.2.1) and then isolating the improvement provided by focus control with a paired-port design (Section 6.2.2), we showed that closed-loop operation materially increases time in focus and suppresses sharpness fluctuations under realistic motion. This is best demonstrated by Supplementary Videos S1-S3. Qualitatively, single-erythrocyte detail is preserved during blinks, drift, and transients, enabling downstream registration,



(a) Representative frames from an autofocus versus fixed-plane experiment across a series of bespoke graticules subjected to a dynamic sinusoidal movement regime.



(b) Representative frames from an autofocus versus fixed-plane experiment across a series of realistic graticules subjected to a dynamic sinusoidal movement regime.

Fig. 6.14 Dynamic autofocus performance across diverse imaging targets.

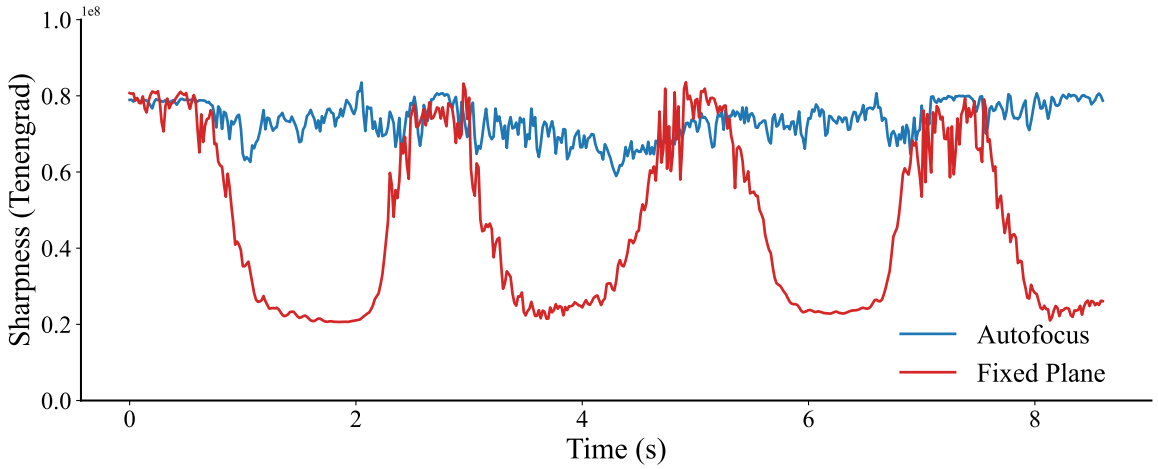


Fig. 6.15 Temporal sharpness traces from a representative dynamic PCB imaging trial. Side-by-side videos of the trials are shown in Supplementary Video S5.

segmentation, and space–time analysis for biomarker extraction (Section 6.2.4). In short, the module resolves a long-standing barrier in HVI by enabling maintenance of single-erythrocyte resolution despite movement artifacts.

We then translated the same autofocus system to a benchtop stereomicroscope and deliberately challenged it with tensile tests on polymer sponges (Section 6.3). Despite a new microscope platform and a converging camera port, the system maintained high sharpness through to the moment of mechanical failure. By comparison, fixed-focus recordings blurred (Fig. 6.11). The Tenengrad traces corroborate sustained clarity following the changing optical distance, and the air-lens position tracked specimen motion smoothly (Fig. 6.13). These results establish platform independence: the module is not limited to the slit-lamp nor even to collimated ports (although they are preferred).

Finally, we surveyed performance across a deliberately heterogeneous panel of imaging targets, from bespoke graticules like grids, lines, and dot fields, to realistic objects like a multilayered microvasculature phantom, a printed-circuit board, and a human hair. Each graticule was subjected to sinusoidal perturbations on the slit-lamp along the optical axis (Section 6.4). In almost all cases, autofocus held focus while fixed-plane controls exhibited periodic collapses in sharpness (Fig. 6.14). The only failure case was near-vertical line gratings with insufficient horizontal texture, which was predictable from first principles. This failure mode is readily mitigated by adding isotropic texture to the scene. That a single controller, score, and actuator preserved high-frequency structure across 65 graticules and four representative objects is strong evidence of target generality.

In sum, these three experiments converge on the same conclusion: the autofocus system achieves video-rate, image-based autofocus that is fast, accurate, and clinically operable

across heterogeneous scenes and optics. It improves the proportion of analyzable frames, stabilises sharpness over time, and preserves detail when motion would otherwise cause defocus. It requires neither calibration nor device-specific alignment, functioning reliably across microscope platforms without major modifications. On this basis, the module justifies the designation used throughout this thesis: a universal, long-range autofocus add-on that generalises across imaging targets and microscope platforms.

Chapter 7

Conclusions

7.1 Summary and discussion

This thesis makes four key contributions:

1. It introduces and validates a universal image-based autofocus device for video microscopy.
2. It delivers a real-time control and acquisition software stack that computes an accurate LoBF from a frame in 1.34 ms, while simultaneously driving the air-lens, performing X–Y image registration, and providing a responsive and full-featured UI.
3. It demonstrates the device's clinical-grade applicability to conjunctival imaging, routinely yielding 40–60 s focused videos at single-erythrocyte resolution (Supplementary Videos S1–S3).
4. It establishes the general applicability of the device by maintaining focus on two slit-lamps, a benchtop stereomicroscope, and across a heterogeneous set of bespoke graticules and real-world targets.

Chapter 2 surveyed the field and made the central problem plain: video-rate microscopy fails when defocus from axial motion is left to post-processing. It showed why radiation-based and existing image-based autofocus struggle outside particular settings, and why learning-based estimators remain too slow and brittle for clinical deployment. From this analysis we distilled specific requirements: to determine both sign and magnitude of defocus from a single exposure; to update at or above 60 Hz; to tolerate illumination asymmetry, sparse texture, and specular reflection; to integrate at a camera port without modifying the internal optics; and to actuate optical power rather than the stage or objective. These

constraints motivated the tilted-sensor plus tunable air-lens architecture developed in the remainder of the thesis. We also set the engineering design goals of speed, accuracy, and clinical operability.

Chapter 3 translated those requirements into a universal optomechanical platform. We formalised the coupling between a tilted sensor and a tunable air-lens, and then validated first-order models with calibration data. Using targeted bench tests, we selected the following components: an Alvium 1800 U-291m imaging camera, a ZWO ASI 120MM-S tilted sensor, an XLA-5-65-1250 refocusing actuator, and an Intel NUC/Linux/C++ processing environment. Relative to earlier prototypes, this architecture moves from short bursts of inaccurate accommodation to sustained video-rate operation, while preserving the integrity of the image beam. A light-tight enclosure completes the transition from benchtop prototype to a deployable, objective-agnostic module.

Chapter 4 provided the software layer that converts capable hardware into a reliable high-speed instrument. A concurrent acquisition–computation–actuation design, a robust location-of-best-focus (LoBF) pipeline, and a gain-scheduled PD controller achieve accurate operation at 60 Hz while recording, rendering, and real-time X–Y image registration proceeds in parallel. These elements are packaged in an open-source, clinically operable desktop application, with an operating manual shown in Appendix A. Effective multi-threaded design preserves GUI responsiveness under load and exposes simple workflows for *Find Focus* and *Hold Focus*. In effect, the sensing-to-accommodation loop is now acquisition-bound rather than computation- or control-bound, and the device is straightforward for non-specialist operators to use.

Chapter 5 moved from architectural intent to empirical characterization under controlled, representative conditions. Closed-loop tests showed rapid, accurate convergence to the set-point across near-step axial disturbances of 0.1–1.0 mm, with settling times $t_s = 50$–142 ms. Optical neutrality was confirmed: the point-spread function broadened only slightly relative to the imaging path (FWHM ratios 1.04 horizontally and 1.08 vertically), and $\mathrm{DOF}_{0.9} \approx 0.07$ was preserved. Crucially, varying the air-lens separation from minimum to maximum induced only a small $\approx 3.4\%$ change in magnification. The autofocus system works across microscope platforms with a working range from 0.5–71.25 mm across eleven objectives and three microscopes (median 9.8 mm), differentiating the device from conventional approaches that are device-specific.

Chapter 6 demonstrated practical value and generality. For *in vivo* conjunctival imaging on a slit-lamp, a paired-port control experiment showed that the autofocus system materially increases time in-focus and suppresses sharpness fluctuations during real conjunctival angiograms. Supplementary Videos S1–S3 demonstrate that single-erythrocyte detail is

preserved despite blinks, drift, and transients. The same autofocus module maintained clarity on an Olympus SZX16 stereomicroscope during polymer-sponge tensile failure, smoothly tracking changes in optical distance to the moment of rupture (Supplementary Video S4). Finally, across a heterogeneous set of 65 graticules and 4 real-world objects, focus was maintained despite sinusoidal axial perturbations (Supplementary Video S5). These three experiments establish clinical value, platform independence, and target generality.

Taken together, Chapters 2–6 substantiate the four contributions enumerated above. They also demonstrate that the system delivers on its design intent, achieving the targeted combination of speed, accuracy, and clinical operability. Nevertheless, as with any novel device, scope remains for further development. Section 7.2 revisits limitations highlighted across the thesis, drawing them together into a concise assessment of performance gaps in the present system. Then, Section 7.3 outlines concrete next steps to guide future work. Finally, Sections 7.3.1–7.3.3 expand the discussion toward clinical translation and future research avenues motivated by this work.

7.2 Limitations

Across the breadth of demonstrations in Chapters 2–6, several limitations with the present device have been observed. Section 6.4 flagged that the edge-based LoBF estimator relies on the presence of some horizontal spatial texture in each frame. By example, while the device can focus on the line graticule shown in Fig. 6.14a, it cannot if that graticule is oriented perfectly vertically. The required texture is minimal, as the sparse-dots target in the same figure illustrates. However, this limitation can surface when imaging a single object that translates laterally across the field of view. In routine use we mitigate this by placing a faint, out-of-focus texture behind or beneath the specimen (e.g., a low-contrast grid or a thin-etched coverslip). More structurally, the constraint arises from the chosen sensor tilt, and could be alleviated by tilting about a different axis or by adopting non-linear tilts, as discussed in the following Section 7.3.

Scene geometry also matters. The system works best when the specimen presents a locally near-flat plane, orthogonal to the optical axis. Strongly tilted or high-depth targets can occasionally bias the LoBF fit, encouraging the controller toward a focus compromise. For *in vivo* slit-lamp imaging, transient partial occlusions from eyebrows and eyelashes in particular can momentarily dominate the sharpness metric, causing the air-lens to “grab” the wrong feature. We reduce such events with temporal gating and outlier rejection in the LoBF algorithm. We also incorporated a sharpness indicator in the GUI, as shown in Figure 4.12,

to help operators orient the microscope orthogonally before starting a recording. But these safeguards do not eliminate failures under severe and consistent occlusions.

A second class of limitations concerns universality across microscopes and mounts. While the actuator–sensor gain is invariant across objectives, small adjustments to controller gains are often required between microscope platforms. We ascribe this to differences in the depth-of-field across microscopes, and to gravity-dependent differences in the Xeryon piezo-actuator. By exposing the relevant controller gain in the settings of the user interface, operators can make minor, one-time adjustments without altering the optical train.

Finally, sensing and acquisition choices introduce practical constraints. The present pipeline is luminance-based and has been validated with a monochrome tilted sensor and imaging camera. We have not yet evaluated colour cameras for use with this autofocus system. Channel-dependent blur could, in principle, degrade the focus estimation, but colour contrast may also supply additional weak texture in feature-poor scenarios. In addition, now that the control loop is acquisition-bound, the tilted sensor's operational limit at 60 Hz becomes rate-limiting. Very rapid axial excursions exceeding 60 Hz can therefore still temporarily evade focus-finding. In sum, these considerations define the envelope of reliable operation for the current implementation and motivate future development steps.

7.3 Future directions

Near-term development at the component-level should consider substitutions that reduce latency and size while keeping the optical train unchanged. Replacing the current Xeryon linear actuator with the recently-released integrated controller model is an obvious step, to remove the existing large control board seen in Figure 3.1. This would significantly reduce the device footprint without affecting operation. In parallel, migrating parts of software and control loop from the acquisition PC to embedded compute (microcontrollers or small FPGAs) could be explored, to make timing more deterministic and simplify the software stack. As highlighted in Section 7.2, however, loop delay is now acquisition-bound. Therefore, the most impactful upgrade is likely a higher frame-rate tilted sensor. A NVIDIA Orin Camera, for example, provides a global shutter and 120 frames per second, which would halve the latency of the autofocus system. A related avenue for development is to attempt sensing in colour, both for the imaging camera and tilted sensor. The current LoBF pipeline could be extended to multi-channel capture to test whether chromatic information supplies additional texture in feature-poor scenarios.

Beyond components, architectural changes to the sensing geometry could improve performance based on system observations. As shown empirically in Section 6.2, 93.2% of lens

instructions were within ± 0.05 cm, or ± 75 pixels, out of an available 320 pixel span. This is partly due to the increased speed of the sensing-actuation loop, showing that the LoBF rarely separates widely from the desired LoBF before the autofocus system accommodates. Based on this finding, a smaller tilt angle, bringing the sensor plane closer to orthogonality with the image beam, would devote more pixels to the actual LoBF range and thereby increase defocus sensitivity. A smaller tilt would also increase geometric correspondence between the images from the tilted sensor and imaging camera, thus aiding sensor fusion efforts. This is discussed further in Section 7.3.2.

In addition, rotating the tilt axis, or using a compound (non-linear) tilt, would mitigate the "horizontal-texture" dependency documented as a limitation in Section 7.2. Figure 7.1 is prepared by Paul Meyer and appears in the patent for this autofocus device, and it illustrates several such orientations. One could also abandon physical sensor tilt in favour of a deliberately distorted intermediate image that maps iso-defocus surfaces to circular (or elliptical) bands on a conventional, untilted sensor. Such a design would preserve the one-frame signed-defocus property while removing the need for a mechanically tilted detector, offering a path toward a thinner, sealed, and potentially more manufacturable module.

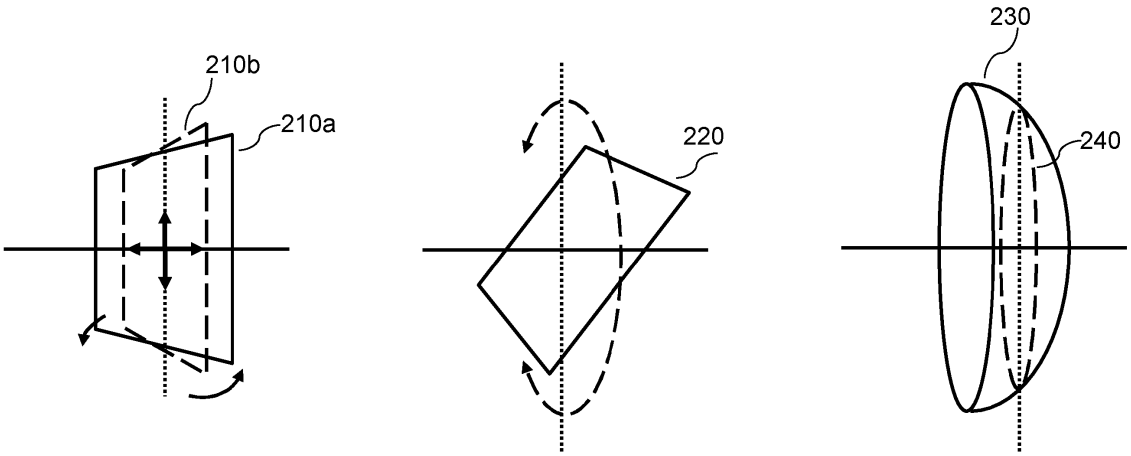


Fig. 7.1 Figure from the patent [87] showing some proposed tilted sensor orientations to provide a more robust sensor for autofocus.

These near-term engineering directions would each improve performance and resolve limitations noted in Section 7.2. However, Section 6.2.2 demonstrated that the device is already clinically valuable, especially for the validated application of conjunctival imaging. Therefore, we next discuss steps for clinical translation, which are already underway.

7.3.1 Clinical translation

Clinical translation has already begun by the development of a pipeline to extract biomarkers from focused videos of the conjunctival microcirculation, as described in Section 6.2.4. The objective is end-to-end, automated extraction of the primary and derived biomarkers summarised in Section 2.4.2: vessel diameter (D), axial velocity (V_a), capillary density, and from these V_c, Q, WSR, and WSS. As mentioned in Section 6.2.4, there is already a bespoke Python-based GUI to extract such information. However, to accelerate development and facilitate transfer to other labs, analysis modules should be re-implemented as plugins to established open-source ecosystems (e.g., Napari or ImageJ), which have well-maintained software and standardised data I/O, metadata capture, frame playback, and quality-control. In parallel, prospective single-centre studies at Addenbrooke's will evaluate inter- and intra-session repeatability, and resolve any operator failure modes that arise. Early clinical studies should aim to provide the normative ranges and reference biomarker characteristics required for clinical reporting.

A natural indication to focus on first, based on Chapter 2's Literature Review, is glaucoma. Research and clinical monitoring of glaucoma benefits directly from high-fidelity, in-focus conjunctival videomicroscopy in two ways. First, hemodynamic assessment of aqueous vessels offers a non-invasive assessment of aqueous outflow (shown in Fig. 2.2). Because HVI-captured aqueous flow appears as a clear core bordered by erythrocytes, velocimetry will depend on proxy metrics inferred from the flanking erythrocyte and viscosity assumptions. Second, surgical outcomes such as trabeculectomy bleb function or stent performance can be evaluated using the same biomarker set used for microcirculation analysis (D, V_a, V_c, WSR, WSS), allowing consistent longitudinal assessment of treatment outcomes. The translational work in glaucoma will focus on establishing test–retest precision for these measures, their sensitivity to intra-ocular pressure perturbations, and their responsiveness to surgery or device intervention.

Diabetes represents a second high-impact application where conjunctival microvascular biomarkers are already known to differ between healthy and diseased populations [15, 69, 62, 3, 51]. Here the emphasis is on reliable quantification of hemorheologic and morphological descriptors that can be mapped to disease stage. A pragmatic clinical protocol is to acquire short, stabilised videos at fixed magnifications and anatomical locations, compute D, V_a, capillary density, and other derived metrics, and aggregate them into interpretable per-eye summaries. While retinal examination will remain essential for assessing diabetic retinopathy, conjunctival imaging may provide a complementary window onto systemic microvascular changes, or potentially serve as an earlier indicator of peripheral vascular dysfunction. The

availability of a clear clinical ground truth (systemic status, fundus-graded retinopathy) would facilitate machine-learning techniques to classify and stratify patients.

7.3.2 Sensor fusion

This work has largely relied on the pairing of the tilted sensor and tunable air-lens for focus accommodation. However, some of the remaining failure modes from section 7.2 fundamentally stem from the difficulties of monocular depth estimation. The natural next step is to fuse the tilted sensor signal with the complementary information available from the imaging camera.

The first operational bridge is already in place. As demonstrated in Supplementary Video S6, the *Get Depths* routine sweeps the LoBF across the tilted sensor and records the LoBF position where each prominent feature is focused on the imaging camera. This effectively links sharpness data from both sensors. After this calibration, the operator may double-click anywhere on the live image to lock focus to that feature, regardless of movement in X, Y, or Z.

Extending this principle into a full fusion pipeline would allow the system to reject misleading pixels and maintain accuracy despite partial occlusion. Practically, this means generating a fixed, calibrated mapping between the two camera planes. Then, by constructing an occlusion or confidence map from the imaging camera and projecting it onto the tilted-sensor view, corrupted regions from eyelids or eyelashes could be dynamically excluded from LoBF estimation. The same mapping would support ROI-based autofocus by weighting the control signal toward a user-selected structure of interest. A smaller tilt angle, as argued earlier, would further tighten geometric correspondence between views. This would improve stereo-mapping and make fusion less brittle to parallax and magnification mismatch.

Fusing the two image streams would resolve several other limitations noted earlier. Texture information from the imaging camera could stabilise focus estimation when the scene lacks horizontal features; frame registration between the two sensors would enable improved autofocus on heavily-curved anatomy. In effect, the two slightly disconjugate views could be merged into a single coherent estimate of depth, yielding autofocus that is both more precise and more tolerant of scene imperfections.

7.3.3 Learning

With a reliable autofocus device, in-focus video acquisition and data collection becomes routine. Learning-based methods can be used to analyse collected videos, but also to improve the instrument itself. Regarding the former, the first goal is to use learning to extract

haemodynamic biomarkers from stabilised clips of the conjunctival microcirculation. Models can be trained to determine the quantities introduced in Section 6.2.4, and those estimators can act as auxiliary heads or regularisers so that disease-risk classifiers remain interpretable. A U-Net model is already used to perform segmentation, and another model used to analyse space-time images, as discussed in Section 6.2.4. With much larger datasets, models can be trained to bypass explicit biomarker extraction and directly distinguish healthy from diseased microcirculations. Even if disease classification can be performed directly, reporting biomarkers will remain essential for clinical interpretability and trust.

Aside from video analysis, learning can also be used to improve the autofocus system itself. The existing *Get Depths* sweep naturally produces supervised pairs: images from both sensors together with the lens command that optimises sharpness for each feature. From these, one can train models for depth prediction, and thus, for focus accommodation. Existing monocular and stereo depth models (e.g. Depth Anything V2, RAFT-Stereo [68]) were briefly trialed for this purpose, but failed due to domain shift and large inter-sensor bias. Purpose-trained models built from microscope-domain datasets are required, and the autofocus system provides an ideal platform for assembling such datasets.

With sufficiently large datasets and effective training, a depth-prediction model would allow the device to perform one-shot autofocus directly on any feature in the image. The operator could double-click on the desired structure, and then the system would instantly infer depth, accommodate, and hold focus as the feature moves. This would collapse the present full-frame workflow to a simpler double-click-to-focus UI.

Ultimately, data-collection, model-building, and device improvement closes a virtuous loop: better focus yields better data, and better data yields better focus. In the limit, the instrument transitions from a clever mechanism into something closer to a pair of trained eyes.

References

[1] Alba, A., Aguilar-Ponce, R. M., Vigueras-Gómez, J. F., and Arce-Santana, E. (2012). Phase correlation based image alignment with subpixel accuracy. In *Proceedings of the 11th Mexican international conference on Advances in Artificial Intelligence*, MICAI, pages 171–182, Berlin, Heidelberg. Springer-Verlag.

[2] Awuah, A., Moore, J. S., Nesbit, M. A., Ruddock, M. W., Brennan, P. F., Mailey, J. A., McNeil, A. J., Jing, M., Finlay, D. D., Trucco, E., Kurth, M. J., Watt, J., Lamont, J. V., Fitzgerald, P., Spence, M. S., McLaughlin, J. A. D., and Moore, T. C. B. (2022). A novel algorithm for cardiovascular screening using conjunctival microcirculatory parameters and blood biomarkers. *Scientific Reports*, 12:6545.

[3] Babenko, B., Mitani, A., Traynis, I., Kitade, N., Singh, P., Maa, A. Y., Cuadros, J., Corrado, G. S., Peng, L., Webster, D. R., Varadarajan, A., Hammel, N., and Liu, Y. (2022). Detection of signs of disease in external photographs of the eyes via deep learning. *Nature Biomedical Engineering*, pages 1–14.

[4] Baker, S. and Matthews, I. (2004). Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*, 56(3):221–255.

[5] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). SURF: Speeded Up Robust Features. In Leonardis, A., Bischof, H., and Pinz, A., editors, *Computer Vision – ECCV 2006*, pages 404–417, Berlin, Heidelberg. Springer.

[6] Born, M. and Wolf, E. (1999). *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, Cambridge, 7 edition.

[7] Brennan, P. F., McNeil, A. J., Jing, M., Awuah, A., Finlay, D. D., Blighe, K., McLaughlin, J. A. D., Wang, R., Moore, J., Nesbit, M. A., Trucco, E., Spence, M. S., and Moore, T. C. B. (2019). Quantitative assessment of the conjunctival microcirculation using a smartphone and slit-lamp biomicroscope. *Microvascular Research*, 126:103907.

[8] Brennan, P. F., McNeil, A. J., Jing, M., Awuah, A., Moore, J. S., Mailey, J., Finlay, D. D., Blighe, K., McLaughlin, J. A. D., Nesbit, M. A., Trucco, E., Moore, T. C. B., and Spence, M. S. (2021). Assessment of the conjunctival microcirculation for patients presenting with acute myocardial infarction compared to healthy controls. *Scientific Reports*, 11(1):7660.

[9] Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698.

[10] Chan, C.-C., Huang, S.-K., and Chen, H. H. (2017). Enhancement of phase detection for autofocus. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 41–45. ISSN: 2381-8549.

[11] Chen, W., Batawi, H. I. M., Alava, J. R., Galor, A., Yuan, J., Sarantopoulos, C. D., McClellan, A. L., Feuer, W. J., Levitt, R. C., and Wang, J. (2017). Bulbar conjunctival microvascular responses in dry eye. *The Ocular Surface*, 15(2):193–201.

[12] Chen, W., Deng, Y., Jiang, H., Wang, J., Zhong, J., Li, S., Peng, L., Wang, B., Yang, R., Zhang, H., Li, M., and Yuan, J. (2018). Microvascular abnormalities in dry eye patients. *Microvascular Research*, 118:155–161.

[13] Cheung, A. T., (Smith) Tomic, M. M., Chen, P. C., Miguelino, E., Li, C.-S., and Devaraj, S. (2009). Correlation of microvascular abnormalities and endothelial dysfunction in Type-1 Diabetes Mellitus (T1DM): A real-time intravital microscopy study. *Clinical Hemorheology and Microcirculation*, 42(4):285–295.

[14] Cheung, A. T. W., Chen, P. C. Y., Larkin, E. C., Duong, P. L., Ramanujam, S., Tablin, F., and Wun, T. (2002). Microvascular abnormalities in sickle cell disease: a computer-assisted intravital microscopy study. *Blood*, 99(11):3999–4005.

[15] Cheung, A. T. W., Ramanujam, S., Greer, D. A., Kumagai, L. F., and Aoki, T. T. (2001). Microvascular Abnormalities in the Bulbar Conjunctiva of Patients with Type 2 Diabetes Mellitus. *Endocrine Practice*, 7(5):358–363.

[16] Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A. W., and Malanda, B. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Research and Clinical Practice*, 138:271–281.

[17] Choi, M., Lee, H., and Lee, H.-e. (2023). Exploring Positional Characteristics of Dual-Pixel Data for Camera Autofocus. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13112–13122, Paris, France. IEEE.

[18] Diehl, P., Brandt, S. R., Morris, M., Gupta, N., and Kaiser, H. (2024). Benchmarking the Parallel 1D Heat Equation Solver in Chapel, Charm++, C++, HPX, Go, Julia, Python, Rust, Swift, and Java. *Springer Nature Switzerland*, 14352:127–138.

[19] Diez-Silva, M., Dao, M., Han, J., Lim, C.-T., and Suresh, S. (2010). Shape and Biomechanical Characteristics of Human Red Blood Cells in Health and Disease. *MRS bulletin / Materials Research Society*, 35(5):382–388.

[20] Dorf, R. C. and Bishop, R. H. (2011). *Modern Control Systems*. Pearson Prentice Hall, Upper Saddle River, NJ, USA, 12 edition.

[21] Du, K., Zhou, D., Zhou, S., Zhang, J., Liu, Q., Bai, X., Liu, Q., Chen, Y., Liu, W., and Kuang, C. (2023). High-accuracy differential autofocus system with an electrically tunable lens. *Optics Letters*, 48(11):2789.

[22] Evangelidis, G. D. and Psarakis, E. Z. (2008). Parametric Image Alignment Using Enhanced Correlation Coefficient Maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865.

[23] Felder, A. E., Mercurio, C., Wanek, J., Ansari, R., and Shahidi, M. (2016). Automated Real-Time Conjunctival Microvasculature Image Stabilization. *IEEE Transactions on Medical Imaging*, 35(7):1670–1675.

[24] Firestone, L., Cook, K., Culp, K., Talsania, N., and Preston Jr., K. (1991). Comparison of autofocus methods for automated microscopy. *Cytometry*, 12(3):195–206.

[25] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395.

[26] Forsberg, D. (2013). *Robust image registration for improved clinical efficiency: using local structure analysis and model-based processing*. Department of Biomedical Engineering, Linköping University, Linköping.

[27] Frangi, A. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A. (1998). Multiscale vessel enhancement filtering. *Springer*, pages 130–137.

[28] Gierl, C., Kondo, T., Voos, H., Kongprawechon, W., and Phoojaruenchanachai, S. (2007). Image Processing Algorithms for an Auto Focus System for Slit Lamp Microscopy. In Blanc-Talon, J., Philips, W., Popescu, D., and Scheunders, P., editors, *Advanced Concepts for Intelligent Vision Systems*, volume 4678, pages 909–919. Springer Berlin Heidelberg, Berlin, Heidelberg.

[29] Goobic, A., Tang, J., and Acton, S. (2005). Image stabilization and registration for tracking cells in the microvasculature. *IEEE Transactions on Biomedical Engineering*, 52(2):287–299.

[30] Groen, F. C., Young, I. T., and Ligthart, G. (1985). A comparison of different focus functions for use in autofocus algorithms. *Cytometry*, 6(2):81–91.

[31] Guizar-Sicairos, M., Thurman, S. T., and Fienup, J. R. (2008). Efficient subpixel image registration algorithms. *Optics Letters*, 33(2):156–158.

[32] Guo, Y.-n., Yi, D.-r., and Li, D.-f. (2019). Rapid autofocus method for optical microscopes using dual photodetectors. In *Proceedings of the 2nd International Conference on Electrical and Electronic Engineering (EEE 2019)*, Advances in Engineering Research, pages 209–216, Paris, France. Atlantis Press.

[33] Hamamatsu (2017). *HCImage – Olympus IX3 Microscope Guide*. Bridgewater, NJ, USA, 1 edition. 18 pp., 2017 Hamamatsu Corporation.

[34] He, D., Cai, D., Zhou, J., Luo, J., and Chen, S.-L. (2019). Adaptive Weighting Depth-variant Deconvolution of Fluorescence Microscopy Images with Convolutional Neural Network. arXiv:1907.03217 [eess].

[35] Herrmann, C., Strong Bowen, R., Wadhwa, N., Garg, R., He, Q., Barron, J. T., and Zabih, R. (2020). Learning to Autofocus. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2227–2236, Seattle, WA, USA. IEEE.

[36] Hessler, M., Arnemann, P.-H., Zamit, F., Seidel, L., Kampmeier, T.-G., Kathöfer, U., Alnawaiseh, M., Tchaichian, S., Rehberg, S., and Ertmer, C. (2019). Monitoring of Conjunctival Microcirculation Reflects Sublingual Microcirculation in Ovine Septic and Hemorrhagic Shock. *Shock*, 51(4):479–486.

[37] Holmes, W. (2022). A Non-Invasive Technique to Observe Microvascular Blood Flow. Master's thesis, Cambridge.

[38] Hsu, W.-Y., Lee, C.-S., Chen, P.-J., Chen, N.-T., Chen, F.-Z., Yu, Z.-R., Kuo, C.-H., and Hwang, C.-H. (2009). Development of the fast astigmatic auto-focus microscope system. *Measurement Science and Technology*, 20(4):045902.

[39] Hu, L., Shi, C., Jiang, H., Shi, Y., Sethi, Z., and Wang, J. (2018). Factors Affecting Microvascular Responses in the Bulbar Conjunctiva in Habitual Contact Lens Wearers. *Investigative Ophthalmology & Visual Science*, 59(10):4108–4114.

[40] Hu, X., Li, Z., Miao, L., Fang, F., Jiang, Z., and Zhang, X. (2023). Measurement Technologies of Light Field Camera: An Overview. *Sensors*, 23(15):6812. Publisher: Multidisciplinary Digital Publishing Institute.

[41] Inoué, S. and Spring, K. R. (1997). *Video Microscopy: The Fundamentals*. Plenum Press, New York, NY, USA, 2 edition.

[42] Ishikawa, A. and Nihoshi, T. (2006). Autofocus system and microscope. U.S. Patent No. 7,071,451 B2. Assignee: Nikon Corporation.

[43] Jerman, T., Pernus, F., Likar, B., and Spiclin, Z. (2016). Enhancement of Vascular Structures in 3D and 2D Angiographic Images. *IEEE transactions on medical imaging*, 35(9):2107–2118.

[44] Jhapate, A. K., Dronawat, R., Saxena, M., and Chourey, R. (2023). Improve the Detection of Retinopathy with Roberts Cross Edge Detection. In Joshi, A., Mahmud, M., and Ragel, R. G., editors, *Information and Communication Technology for Competitive Strategies (ICTCS 2021)*, Lecture Notes in Networks and Systems, pages 475–483, Singapore. Springer Nature.

[45] Jiang, H., Ye, Y., DeBuc, D. C., Lam, B. L., Rundek, T., Tao, A., Shao, Y., and Wang, J. (2013). Human conjunctival microvasculature assessed with a retinal function imager (RFI). *Microvascular Research*, 85:134–137.

[46] Jiang, H., Zhong, J., DeBuc, D. C., Tao, A., Xu, Z., Lam, B. L., Liu, C., and Wang, J. (2014). Functional slit lamp biomicroscopy for imaging bulbar conjunctival microvasculature in contact lens wearers. *Microvascular Research*, 92:62–71.

[47] Jin, L., Tang, Y., Wu, Y., Coole, J. B., Tan, M. T., Zhao, X., Badaoui, H., Robinson, J. T., Williams, M. D., Gillenwater, A. M., Richards-Kortum, R. R., and Veeraraghavan, A. (2020). Deep learning extended depth-of-field microscope for fast and slide-free histology. *Proceedings of the National Academy of Sciences*, 117(52):33051–33060.

[48] Johnstone, M., Martin, E., and Jamil, A. (2011). Pulsatile flow into the aqueous veins: Manifestations in normal and glaucomatous eyes. *Experimental eye research*, 92(5):318–327.

[49] Johnstone, M., Xin, C., Tan, J., Martin, E., Wen, J., and Wang, R. K. (2021). Aqueous outflow regulation – 21st century concepts. *Progress in Retinal and Eye Research*, 83:100917.

[50] Khansari, M. M., Wanek, J., Felder, A. E., Camardo, N., and Shahidi, M. (2016). Automated Assessment of Hemodynamics in the Conjunctival Microvasculature Network. *IEEE Transactions on Medical Imaging*, 35(2):605–611.

[51] Khansari, M. M., Wanek, J., Tan, M., Joslin, C. E., Kresovich, J. K., Camardo, N., Blair, N. P., and Shahidi, M. (2017). Assessment of Conjunctival Microvascular Hemodynamics in Stages of Diabetic Microvasculopathy. *Scientific Reports*, 7(1):45916.

[52] Khatib, T. Z., Meyer, P. A., Lusthaus, J., Manyakin, I., Mushtaq, Y., and Martin, K. R. (2019). Hemoglobin Video Imaging Provides Novel In Vivo High-Resolution Imaging and Quantification of Human Aqueous Outflow in Patients with Glaucoma. *Ophthalmology. Glaucoma*, 2(5):327–335.

[53] Knapper, J., Collins, J. T., Stirling, J., McDermott, S., Wadsworth, W., and Bowman, R. W. (2022). Fast, high-precision autofocus on a motorised microscope: Automating blood sample imaging on the OpenFlexure Microscope. *Journal of Microscopy*, 285(1):29–39.

[54] Konathala, L. (2024). Analysing Microvascular Blood Flow with Machine Learning Computer Vision. Undergraduate thesis, University of Cambridge.

[55] Korneeva, N. V., Syrotin, B. Z., Bondar, N. V., and Rudman, V. Y. (2019). The relationship of microcirculation parameters of the bulbar conjunctiva and coronary angiography in patients with coronary artery disease. *Russian Journal of Cardiology*, 0(2):44–52.

[56] Koutsiaris, A. G. (2005). Volume flow estimation in the precapillary mesenteric microvasculature in vivo and the principle of constant pressure gradient. *Biorheology*, 42(6):479–491.

[57] Koutsiaris, A. G., Tachmitzi, S. V., Papavasileiou, P., Batis, N., Kotoula, M. G., Giannoukas, A. D., and Tsironi, E. (2010). Blood velocity pulse quantification in the human conjunctival pre-capillary arterioles. *Microvascular Research*, 80(2):202–208.

[58] Kvernebo, A. K., Miyamoto, T., Drolsum, L., Moe, M. C., Måsøy, S.-E., Sunagawa, G., Dessoffy, R., Karimov, J. H., Fukamachi, K., and Kvernebo, K. (2022). Ocular surface microcirculation is better preserved with pulsatile versus continuous flow during cardiopulmonary bypass—An experimental pilot. *Artificial Organs*, 46(5):786–793.

[59] Kvernebo, A. K., Miyamoto, T., Sporastøyl, A. H., Wikslund, L. K., Måsøy, S.-E., Drolsum, L., Moe, M. C., Salerud, G., Fukamachi, K., and Kvernebo, K. (2020). Quantification of ocular surface microcirculation by computer assisted video microscopy and diffuse reflectance spectroscopy. *Experimental Eye Research*, 201:108312.

[60] Lachica, E. (2006). Keeping Life in Focus New Systems Prevent Z-axis Drift in Time Lapse Studies. *Microscopy Today*, 14(4):42–46.

[61] Laine, R. F., Heil, H. S., Coelho, S., Nixon-Abell, J., Jimenez, A., Wiesner, T., Martínez, D., Galgani, T., Régnier, L., Stubb, A., Follain, G., Webster, S., Goyette, J., Dauphin, A., Salles, A., Culley, S., Jacquemet, G., Hajj, B., Leterrier, C., and Henriques, R. (2023). High-fidelity 3D live-cell nanoscopy through data-driven enhanced super-resolution radial fluctuation. *Nature Methods*, 20(12):1949–1956. Publisher: Nature Publishing Group.

[62] Li, X., Xia, C., Li, X., Wei, S., Zhou, S., Yu, X., Gao, J., Cao, Y., and Zhang, H. (2022). Identifying diabetes from conjunctival images using a novel hierarchical multi-task network. *Scientific Reports*, 12(1):264.

[63] Liao, J., Bian, L., Bian, Z., Zhang, Z., Patel, C., Hoshino, K., Eldar, Y. C., and Zheng, G. (2016). Single-frame rapid autofocusing for brightfield and fluorescence whole slide imaging. *Biomedical Optics Express*, 7(11):4763–4768.

[64] Liao, J., Chen, X., Ding, G., Dong, P., Ye, H., Wang, H., Zhang, Y., and Yao, J. (2021a). Deep learning-based single-shot autofocus method for digital microscopy. *Biomedical Optics Express*, 13(1):314–327.

[65] Liao, Y., Xiong, Y., and Yang, Y. (2021b). An Auto-Focus Method of Microscope for the Surface Structure of Transparent Materials under Transmission Illumination. *Sensors*, 21(7):2487. Publisher: Multidisciplinary Digital Publishing Institute.

[66] Lightley, J., Kumar, S., Lim, M. Q., Garcia, E., Görlitz, F., Alexandrov, Y., Parrado, T., Hollick, C., Steele, E., Roßmann, K., Graham, J., Broichhagen, J., McNeish, I. A., Roufosse, C. A., Neil, M. A. A., Dunsby, C., and French, P. M. W. (2023). openFrame: A modular, sustainable, open microscopy platform with single-shot, dual-axis optical autofocus module providing high precision and long range of operation. *Journal of Microscopy*, 292(2):64–77.

[67] Lim, Y., Cha, M. C., and Chang, J. Y. (2015). Compressible and monolithic microporous polymer sponges prepared via one-pot synthesis. *Scientific Reports*, 5(1):15957. Publisher: Nature Publishing Group.

[68] Lipson, L., Teed, Z., and Deng, J. (2021). RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching. arXiv:2109.07547 [cs].

[69] Liu, Z., Jiang, H., Townsend, J. H., and Wang, J. (2020). Improved conjunctival microcirculation in diabetic retinopathy patients with MTHFR polymorphisms after Ocufolin™ Administration. *Microvascular Research*, 132:104066.

[70] Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110.

[71] Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2*, IJCAI'81, pages 674–679, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[72] Lusthaus, J. A., Khatib, T. Z., Meyer, P. A. R., McCluskey, P., and Martin, K. R. (2021). Aqueous outflow imaging techniques and what they tell us about intraocular pressure regulation. *Eye*, 35(1):216–235.

[73] Lusthaus, J. A., Meyer, P. A. R., Khatib, T. Z., and Martin, K. R. (2020). The Effects of Trabecular Bypass Surgery on Conventional Aqueous Outflow, Visualized by Hemoglobin Video Imaging. *Journal of Glaucoma*, 29(8):656–665.

[74] Lynch, T. (2025). Videos Included with the PhD Thesis, "A Universal Image-Based Autofocus Device for Video Microscopy". Zenodo. DOI: 10.5281/zenodo.17592699. Available at: https://zenodo.org/records/17592699.

[75] Ma, D. H.-K., Chen, J.-K., Zhang, F., Lin, K.-Y., Yao, J.-Y., and Yu, J.-S. (2006). Regulation of corneal angiogenesis in limbal stem cell deficiency. *Progress in Retinal and Eye Research*, 25(6):563–590.

[76] Markwirth, A., Lachetta, M., Mönkemöller, V., Heintzmann, R., Hübner, W., Huser, T., and Müller, M. (2019). Video-rate multi-color structured illumination microscopy with simultaneous real-time reconstruction. *Nature Communications*, 10(1):4315. Publisher: Nature Publishing Group.

[77] Mateos-Pérez, J. M., Redondo, R., Nava, R., Valdiviezo, J. C., Cristóbal, G., Escalante-Ramírez, B., Ruiz-Serrano, M. J., Pascau, J., and Desco, M. (2012). Comparative evaluation of autofocus algorithms for a real-time system for automatic detection of Mycobacterium tuberculosis. *Cytometry Part A*, 81A(3):213–221.

[78] McKay, R. R., Baxi, V. A., and Montalto, M. C. (2011). The accuracy of dynamic predictive autofocusing for whole slide imaging. *Journal of Pathology Informatics*, 2:38.

[79] Meyer, P. A. R. (1989). The circulation of the human limbus. *Eye*, 3(2):121–127.

[80] Meyer, P. A. R. (2018). Re-orchestration of blood flow by micro-circulations. *Eye*, 32(2):222–229.

[81] Meyer, P. A. R., Das, T., Ghadiri, N., Murthy, R., and Theodoropoulou, S. (2019). Clinical pathophysiology of thyroid eye disease: The Cone Model. *Eye*, 33(2):244–253.

[82] Meyer, P. A. R. and Watson, P. G. (1987). Low dose fluorescein angiography of the conjunctiva and episclera. *British Journal of Ophthalmology*, 71(1):2–10.

[83] Montalto, M. C., McKay, R. R., and Filkins, R. J. (2011). Autofocus methods of whole slide imaging systems and the introduction of a second-generation independent dual sensor scanning method. *Journal of Pathology Informatics*, 2:44.

[84] Mäki-Petäjä, K. M., McGeoch, A., Yang, L. L., Hubsch, A., McEniery, C. M., Meyer, P. A., Mir, F., Gajendragadkar, P., Ramenatte, N., Anandappa, G., Santos Franco, S., Bond, S. J., Schönlieb, C.-B., Boink, Y., Brune, C., Wilkinson, I. B., Jodrell, D. I., and Cheriyan, J. (2021). Mechanisms Underlying Vascular Endothelial Growth Factor Receptor Inhibition–Induced Hypertension: The HYPAZ Trial. *Hypertension*, 77(5):1591–1599.

[85] Niederauer, C., Seynen, M., Zomerdijk, J., Kamp, M., and Ganzinger, K. A. (2023). The K2: Open-source simultaneous triple-color TIRF microscope for live-cell and single-molecule imaging. *HardwareX*, 13:e00404.

[86] Patel, N. R., Duong, S., Mirbod, P., Xu, J., and Gaynes, B. I. (2022). A meta-analysis of variability in conjunctival microvascular hemorheology metrics. *Microvascular Research*, 142:104340.

[87] Paul, M., Savin, T., and Lynch, T. (2025). Autofocus during video-microscopy. Patent GB2632324A, filed by Cambridge Enterprise Ltd. Available at https://patents.google.com/patent/GB2632324A/en.

[88] Pawley, J. B., editor (2006). *Handbook of Biological Confocal Microscopy*. Springer Science and Business Media, New York, NY, USA, 3 edition.

[89] Pickering, S., Batra, R., Merrick, B., Snell, L. B., Nebbia, G., Douthwaite, S., Reid, F., Patel, A., Ik, M. T. K., Patel, B., Charalampous, T., Alcolea-Medina, A., Lista, M. J., Cliff, P. R., Cunningham, E., Mullen, J., Doores, K. J., Edgeworth, J. D., Malim, M. H., Neil, S. J. D., and Galão, R. P. (2021). Comparative performance of SARS-CoV-2 lateral flow antigen tests and association with detection of infectious virus in clinical specimens: a single-centre laboratory evaluation study. *The Lancet Microbe*, 2(9):e461–e471.

[90] Pinkard, H., Phillips, Z., Babakhani, A., Fletcher, D. A., and Waller, L. (2019). Deep learning for single-shot autofocus microscopy. *Optica*, 6(6):794.

[91] Rahmani, A., Cox, T., Achary, A. T. A., and Ponjavic, A. (2024). Astigmatism-based active focus stabilisation with universal objective lens compatibility, extended operating range and nanometer precision. *Optics Express*, 32(8):13331–13341. Publisher: Optica Publishing Group.

[92] Rosten, E. and Drummond, T. (2006). Machine Learning for High-Speed Corner Detection. In Leonardis, A., Bischof, H., and Pinz, A., editors, *Computer Vision – ECCV 2006*, pages 430–443, Berlin, Heidelberg. Springer.

[93] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*, pages 2564–2571. ISSN: 2380-7504.

[94] Rudolph, T., Wyder, S., De Zanet, S., Wolf, S., and Kowal, J. (2013). A fast auto-focus for digital slit lamp imaging. *Investigative Ophthalmology & Visual Science*, 54(15):1516.

[95] Santos, A., Ortiz De Solórzano, C., Vaquero, J. J., Peña, J. M., Malpica, N., and Del Pozo, F. (1997). Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of Microscopy*, 188(3):264–272.

[96] Shah, M., Mishra, S., Sarkar, M., and Rout, C. (2017). Identification of robust focus measure functions for the automated capturing of focused images from Ziehl–Neelsen stained sputum smear microscopy slide. *Cytometry Part A*, 91(8):800–809.

[97] Shen, F., Hodgson, L., and Hahn, K. (2006). Digital Autofocus Methods for Automated Microscopy. In *Methods in Enzymology*, volume 414, pages 620–632. Academic Press. ISSN: 0076-6879.

[98] Shi, C., Jiang, H., Gameiro, G. R., and Wang, J. (2019). Microcirculation in the conjunctiva and retina in healthy subjects. *Eye and Vision*, 6(1):11.

[99] Shi, J. and Tomasi (1994). Good features to track. In *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600. ISSN: 1063-6919.

[100] Shih, L. (2007). Autofocus survey: a comparison of algorithms. In *Digital Photography III*, volume 6502, pages 90–100. SPIE.

[101] Shu, X., Wang, J., and Hu, L. (2019). A review of functional slit lamp biomicroscopy. *Eye and Vision*, 6(1):15.

[102] Silfies, J. (2025). The Nikon Perfect Focus System (PFS). Technical manual, Nikon Microscopy. Available at: https://www.microscopyu.com/tutorials/the-nikon-perfect-focus-system-pfs [Accessed 24 Sep 2025].

[103] Silvestri, L., Müllenbroich, M. C., Costantini, I., Di Giovanna, A. P., Mazzamuto, G., Franceschini, A., Kutra, D., Kreshuk, A., Checcucci, C., Toresano, L. O., Frasconi, P., Sacconi, L., and Pavone, F. S. (2021). Universal autofocus for quantitative volumetric microscopy of whole mouse brains. *Nature Methods*, 18(8):953–958. Publisher: Nature Publishing Group.

[104] Smith, M. M., Chen, P. C. Y., Li, C.-S., Ramanujam, S., and Cheung, A. T. W. (2009). Whole blood viscosity and microvascular abnormalities in Alzheimer's Disease. *Clinical Hemorheology and Microcirculation*, 41(4):229–239.

[105] Sullivan, J. M., Prewitt, R. L., and Josephs, J. A. (1983). Attenuation of the microcirculation in young patients with high-output borderline hypertension. *Hypertension (Dallas, Tex.: 1979)*, 5(6):844–851.

[106] Sun, Y., Chen, A., Zou, M., Zhang, Y., Jin, L., Li, Y., Zheng, D., Jin, G., and Congdon, N. (2022). Time trends, associations and prevalence of blindness and vision loss due to glaucoma: an analysis of observational data from the Global Burden of Disease Study 2017. *BMJ open*, 12(1):e053805.

[107] Sun, Y., Duthaler, S., and Nelson, B. J. (2004). Autofocusing in computer microscopy: selecting the optimal focus algorithm. *Microscopy Research and Technique*, 65(3):139–149.

[108] Suzuki, M., Suzuki, Y., Komori, R., Orii, Y., Arimura, S., Iwasaki, K., Takamura, Y., and Inatani, M. (2022). Aqueous column changes in the episcleral veins after the instillation of ripasudil versus latanoprost: a randomized, double-blind, crossover clinical trial. *Scientific Reports*, 12(1):15255.

[109] Szeliski, R. (2007). Image Alignment and Stitching: A Tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1):1–104.

[110] Thevenaz, P., Ruttimann, U., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Transactions on Image Processing*, 7(1):27–41.

[111] Tiwari, A., Myeong, J., Hashemiaghdam, A., Stunault, M. I., Zhang, H., Niu, X., Laramie, M. A., Sponagel, J., Shriver, L. P., Patti, G. J., Klyachko, V. A., and Ashrafi, G.

(2024). Mitochondrial pyruvate transport regulates presynaptic metabolism and neurotransmission. *Science Advances*, 10(46):eadp7423. Publisher: American Association for the Advancement of Science.

[112] Tuma, R. F., Duran, W. N., and Ley, K. (2011). *Microcirculation*. Academic Press, San Diego, CA, USA.

[113] Van Steenbergen, V., Lavoie-Cardinal, F., Kazwiny, Y., Decet, M., Martens, T., Verstreken, P., Boesmans, W., De Koninck, P., and Vanden Berghe, P. (2022). Nanopositioning and tubulin conformation contribute to axonal transport regulation of mitochondria along microtubules. *Proceedings of the National Academy of Sciences*, 119(45):e2203499119.

[114] van Zijderveld, R., Ince, C., and Schlingemann, R. O. (2014). Orthogonal polarization spectral imaging of conjunctival microcirculation. *Graefe's Archive for Clinical and Experimental Ophthalmology*.

[115] Westphal, P. and Mitzkus, R. (2014). Autofocus method for microscope and microscope with autofocus device. Patent US20140168404A1, Carl Zeiss Microscopy GmbH. Available at: https://patents.google.com/patent/US20140168404A1/en [Accessed 25 Sep 2025].

[116] Xu, Z., Jiang, H., Tao, A., Wu, S., Yan, W., Yuan, J., Liu, C., DeBuc, D. C., and Wang, J. (2015). Measurement variability of the bulbar conjunctival microvasculature in healthy subjects using functional slit lamp biomicroscopy (FSLB). *Microvascular Research*, 101:15–19.

[117] Yazdanfar, S., Kenny, K. B., Tasimi, K., Corwin, A. D., Dixon, E. L., and Filkins, R. J. (2008). Simple and robust image-based autofocusing for digital microscopy. *Optics Express*, 16(12):8670–8677. Publisher: Optica Publishing Group.

[118] Yun, Z., Xu, Q., Wang, G., Jin, S., Lin, G., Feng, Q., and Yuan, J. (2022). EVA: Fully automatic hemodynamics assessment system for the bulbar conjunctival microvascular network. *Computer Methods and Programs in Biomedicine*, 216:106631.

[119] Zhang, H. and Yao, J. (2021). Automatic Focusing Method of Microscopes Based on Image Processing. *Mathematical Problems in Engineering*, 2021(1):8243072.

[120] Zhang, J., Fu, Y.-F., Shen, H., Liu, Q., Sun, L.-N., and Chen, L.-G. (2024). Precision autofocus in optical microscopy with liquid lenses controlled by deep reinforcement learning. *Microsystems & Nanoengineering*, 10(1):201.

[121] Zhang, J. and Zhang, T. (2021). Focusing Algorithm of Automatic Control Microscope Based on Digital Image Processing. *Journal of Sensors*, 2021(1):5643054.

[122] Zhao, S., Yang, Z., Sun, P., Wu, X., Tang, W., Shao, F., and Tang, Z. (2020). Conjunctival microcirculation is associated with cerebral cortex microcirculation in post-resuscitation mild hypothermia: A rat model. *Microcirculation*, 27(3):e12604.

[123] Zhu, Y., Makita, S., Fukutake, N., and Yasuno, Y. (2025). Theoretical analysis of performance limitation of computational refocusing in optical coherence tomography. arXiv:2501.13874 [physics].

[124] Ziegler, J. G. and Nichols, N. B. (1942). Optimum settings for automatic controllers. *Transactions of the ASME*, 64(11):759–765.

[125] Zitová, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000.

Appendix A

Operating Manual

Quick start: acquire a new recording

1. Plug in the autofocus device and start the `autofocus-anything` app. The lens will perform a brief $\sim$1.5 s homing sequence on start-up.
2. Align the microscope near the desired focal plane and adjust camera settings until the image is well lit.
3. Click **Find Focus**. The device will automatically find and continuously hold the focal plane that maximizes sharpness on the imaging camera.
    - Optionally use the **Best-Focal Plane** slider to maintain a slightly deeper or shallower plane while staying locked in-focus.
4. Click **Record**. By default, at 60 fps the software captures a 20-second recording.
5. When the recording finishes, click **Save**. Enter a folder name; the software will create that folder (if needed) and save the full-resolution image frames in sequence.

Tip. If you prefer to hold the exact plane currently visible (without searching for maximum sharpness first), use **Hold Focus** instead of **Find Focus**.

Load and view a past recording

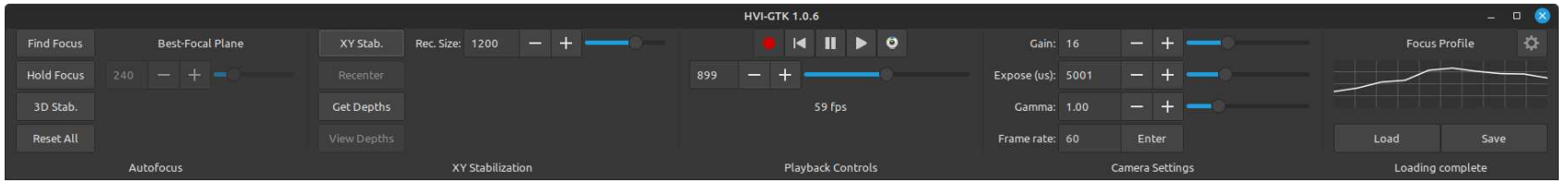
1. Open the `autofocus-anything` app and click **Load**.
2. Select the folder containing the recorded frames and click **Open**. The progress indicator below **Load/Save** shows loading status (typically $\sim$2 s for a 60-s video).

3. Use the playback buttons and progress slider to review the recording.

Tip. Use the **XY Stab.** button to stabilize the image during playback or while viewing a live feed. The software offsets each frame to compensate for lateral motion.

Tip. You can zoom with the mouse wheel or pan by clicking and dragging.

A full feature-by-feature breakdown is shown in Fig. A.1.



Find Focus — Moves the lens to maximize sharpness across the image, and then continuously holds that focal plane.

Hold Focus — Moves the lens to continuously hold the *current* focal plane.

3D Stab. — Runs **Hold Focus** and **XY Stab.** simultaneously.

Reset All — Returns lens to home position, and resets any frame offset.

Best-Focal Plane (240) — Becomes active when **Hold Focus** or **Find Focus** are active. Sliding to left shifts the focal plane deeper, and to the right shifts it shallower.

XY Stab. — Stabilizes frames on features in real-time by applying an X-Y offset.

Recenter — Resets the X-Y offset, recentering the frames.

Get Depths — Sweeps the lens and records where each feature is sharpest, creating a depth map. Allows double-click to autofocus on a specific feature.

View Depths — Toggles overlaying the depth map on the scene.

Rec. Size: 1200 — Recording size in # of frames.

Starts or continues a recording.

When rewatching a recording, returns to the start.

Pauses a recording or pauses playback of a past recording.

Plays playback of a recording.

Returns to the live camera feed from a loaded recording.

Focus Profile — Shows a sharpness profile across the image horizontal, used to ensure the microscope is orthogonal.

Load / Save — For loading and saving recordings.

Loading complete — **Load**/**Save** progress indicator.

Provides user access to various settings, including:

1. **Controller gain**: Lower for a slower, more stable response.
2. **Find Focus** home: the location on the tilted sensor that corresponds with a sharp imaging camera, typically near the center. Used when finding focus.

Fig. A.1 Annotated overview of the autofocus software graphical user interface (GUI). Reproduced with minor modifications from Fig. 4.12 for convenience.